

# Deep and Transfer Learning Approaches for Complex Data Analysis in the Industry 4.0 Era

Lead Guest Editor: Yuanpeng Zhang

Guest Editors: Shan Zhong, Khin wee Lai, and Pengjiang Qian





---

# **Deep and Transfer Learning Approaches for Complex Data Analysis in the Industry 4.0 Era**

Wireless Communications and Mobile Computing

---

**Deep and Transfer Learning Approaches  
for Complex Data Analysis in the  
Industry 4.0 Era**

Lead Guest Editor: Yuanpeng Zhang

Guest Editors: Shan Zhong, Khin wee Lai, and  
Pengjiang Qian




---



Copyright © 2023 Hindawi Limited. All rights reserved.

This is a special issue published in “Wireless Communications and Mobile Computing.” All articles are open access articles distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

# Chief Editor

Zhipeng Cai , USA

## Associate Editors

Ke Guan , China  
Jaime Lloret , Spain  
Maode Ma , Singapore

## Academic Editors

Muhammad Inam Abbasi, Malaysia  
Ghufran Ahmed , Pakistan  
Hamza Mohammed Ridha Al-Khafaji ,  
Iraq  
Abdullah Alamoodi , Malaysia  
Marica Amadeo, Italy  
Sandhya Aneja, USA  
Mohd Dilshad Ansari, India  
Eva Antonino-Daviu , Spain  
Mehmet Emin Aydin, United Kingdom  
Parameshchhari B. D. , India  
Kalapaveen Bagadi , India  
Ashish Bagwari , India  
Dr. Abdul Basit , Pakistan  
Alessandro Bazzi , Italy  
Zdenek Becvar , Czech Republic  
Nabil Benamar , Morocco  
Olivier Berder, France  
Petros S. Bithas, Greece  
Dario Bruneo , Italy  
Jun Cai, Canada  
Xuesong Cai, Denmark  
Gerardo Canfora , Italy  
Rolando Carrasco, United Kingdom  
Vicente Casares-Giner , Spain  
Brijesh Chaurasia, India  
Lin Chen , France  
Xianfu Chen , Finland  
Hui Cheng , United Kingdom  
Hsin-Hung Cho, Taiwan  
Ernestina Cianca , Italy  
Marta Cimitile , Italy  
Riccardo Colella , Italy  
Mario Collotta , Italy  
Massimo Condoluci , Sweden  
Antonino Crivello , Italy  
Antonio De Domenico , France  
Floriano De Rango , Italy

Antonio De la Oliva , Spain  
Margot Deruyck, Belgium  
Liang Dong , USA  
Praveen Kumar Donta, Austria  
Zhuojun Duan, USA  
Mohammed El-Hajjar , United Kingdom  
Oscar Esparza , Spain  
Maria Fazio , Italy  
Mauro Femminella , Italy  
Manuel Fernandez-Veiga , Spain  
Gianluigi Ferrari , Italy  
Luca Foschini , Italy  
Alexandros G. Fragkiadakis , Greece  
Ivan Ganchev , Bulgaria  
Óscar García, Spain  
Manuel García Sánchez , Spain  
L. J. García Villalba , Spain  
Miguel Garcia-Pineda , Spain  
Piedad Garrido , Spain  
Michele Girolami, Italy  
Mariusz Glabowski , Poland  
Carles Gomez , Spain  
Antonio Guerrieri , Italy  
Barbara Guidi , Italy  
Rami Hamdi, Qatar  
Tao Han, USA  
Sherief Hashima , Egypt  
Mahmoud Hassaballah , Egypt  
Yejun He , China  
Yixin He, China  
Andrej Hrovat , Slovenia  
Chunqiang Hu , China  
Xuexian Hu , China  
Zhenghua Huang , China  
Xiaohong Jiang , Japan  
Vicente Julian , Spain  
Rajesh Kaluri , India  
Dimitrios Katsaros, Greece  
Muhammad Asghar Khan, Pakistan  
Rahim Khan , Pakistan  
Ahmed Khattab, Egypt  
Hasan Ali Khattak, Pakistan  
Mario Kolberg , United Kingdom  
Meet Kumari, India  
Wen-Cheng Lai , Taiwan

Jose M. Lanza-Gutierrez, Spain  
Pavlos I. Lazaridis , United Kingdom  
Kim-Hung Le , Vietnam  
Tuan Anh Le , United Kingdom  
Xianfu Lei, China  
Jianfeng Li , China  
Xiangxue Li , China  
Yaguang Lin , China  
Zhi Lin , China  
Liu Liu , China  
Mingqian Liu , China  
Zhi Liu, Japan  
Miguel López-Benítez , United Kingdom  
Chuanwen Luo , China  
Lu Lv, China  
Basem M. ElHalawany , Egypt  
Imadeldin Mahgoub , USA  
Rajesh Manoharan , India  
Davide Mattera , Italy  
Michael McGuire , Canada  
Weizhi Meng , Denmark  
Klaus Moessner , United Kingdom  
Simone Morosi , Italy  
Amrit Mukherjee, Czech Republic  
Shahid Mumtaz , Portugal  
Giovanni Nardini , Italy  
Tuan M. Nguyen , Vietnam  
Petros Nicolaitidis , Greece  
Rajendran Parthiban , Malaysia  
Giovanni Pau , Italy  
Matteo Petracca , Italy  
Marco Picone , Italy  
Daniele Pinchera , Italy  
Giuseppe Piro , Italy  
Javier Prieto , Spain  
Umair Rafique, Finland  
Maheswar Rajagopal , India  
Sujan Rajbhandari , United Kingdom  
Rajib Rana, Australia  
Luca Reggiani , Italy  
Daniel G. Reina , Spain  
Bo Rong , Canada  
Mangal Sain , Republic of Korea  
Praneet Saurabh , India

Hans Schotten, Germany  
Patrick Seeling , USA  
Muhammad Shafiq , China  
Zaffar Ahmed Shaikh , Pakistan  
Vishal Sharma , United Kingdom  
Kaize Shi , Australia  
Chakchai So-In, Thailand  
Enrique Stevens-Navarro , Mexico  
Sangeetha Subbaraj , India  
Tien-Wen Sung, Taiwan  
Suhua Tang , Japan  
Pan Tang , China  
Pierre-Martin Tardif , Canada  
Sreenath Reddy Thummaluru, India  
Tran Trung Duy , Vietnam  
Fan-Hsun Tseng, Taiwan  
S Velliangiri , India  
Quoc-Tuan Vien , United Kingdom  
Enrico M. Vitucci , Italy  
Shaohua Wan , China  
Dawei Wang, China  
Huaqun Wang , China  
Pengfei Wang , China  
Dapeng Wu , China  
Huaming Wu , China  
Ding Xu , China  
YAN YAO , China  
Jie Yang, USA  
Long Yang , China  
Qiang Ye , Canada  
Changyan Yi , China  
Ya-Ju Yu , Taiwan  
Marat V. Yuldashev , Finland  
Sherali Zeadally, USA  
Hong-Hai Zhang, USA  
Jiliang Zhang, China  
Lei Zhang, Spain  
Wence Zhang , China  
Yushu Zhang, China  
Kechen Zheng, China  
Fuhui Zhou , USA  
Meiling Zhu, United Kingdom  
Zhengyu Zhu , China

# Contents

**Retracted: Construction and Analysis of Green Investment Risk Evaluation Index System Based on Information Entropy Fuzzy Hierarchical Analysis Model**

Wireless Communications and Mobile Computing  
Retraction (1 page), Article ID 9828050, Volume 2023 (2023)

**Retracted: Influencing Factors of Microlecture on the Teaching Effect of Ideological and Political Courses in Colleges**

Wireless Communications and Mobile Computing  
Retraction (1 page), Article ID 9760597, Volume 2023 (2023)


**Retracted: Deep Convolutional Neural Network and Weighted Bayesian Model for Evaluation of College Foreign Language Multimedia Teaching**

Wireless Communications and Mobile Computing  
Retraction (1 page), Article ID 9761817, Volume 2023 (2023)



**A Framework for Identification and Classification of IoT Devices for Security Analysis in Heterogeneous Network**

Hafiz Muhammad Zahid, Yasir Saleem, Faisal Hayat, Farrukh Zeeshan Khan , Roobaea Alroobaea , Fahad Almansour, Muneer Ahmad , and Ihsan Ali   
Research Article (16 pages), Article ID 8806184, Volume 2022 (2022)


**Application of High-Resolution Remote Sensing Image for Individual Tree Identification of Pinus sylvestris and Pinus tabulaeformis**

Hong Li and Wunian Yang   
Research Article (11 pages), Article ID 7672762, Volume 2021 (2021)

**Multiscale Bidirectional Input Convolutional and Deep Neural Network for Human Activity Recognition**

Yishu Qiu, Lanliang Lin , Lvqing Yang , Dingzhao Li, Runhan Song, Gengchen Xu, and Shaoqin Shen  
Research Article (9 pages), Article ID 7374177, Volume 2021 (2021)


**Research on the Influencing Factors of Film Consumption and Box Office Forecast in the Digital Era: Based on the Perspective of Machine Learning and Model Integration**

Qi He and Bin Hu   
Research Article (10 pages), Article ID 6094924, Volume 2021 (2021)


**Big Data and Deep Learning-Based Video Classification Model for Sports**

Lin Wang, Haiyan Zhang , and Guoliang Yuan  
Research Article (11 pages), Article ID 1140611, Volume 2021 (2021)

**Innovation and Practice of Music Education Paths in Universities under the Popularity of 5G Network**


Haoyu Cao   
Research Article (11 pages), Article ID 3570412, Volume 2021 (2021)

**Research and Implementation of the Sports Analysis System Based on 3D Image Technology**

Hongwei Wang, Jie Gao, and Jingjing Liu 


Research Article (11 pages), Article ID 4266417, Volume 2021 (2021)

**Research and Development of Inventory Management and Human Resource Management in ERP**

Bo Zhao  and Chunlei Tu


Research Article (12 pages), Article ID 3132062, Volume 2021 (2021)

**[Retracted] Influencing Factors of Microlecture on the Teaching Effect of Ideological and Political Courses in Colleges**

Yuqian Jin 

Research Article (7 pages), Article ID 6309221, Volume 2021 (2021)

**Construction and Simulation of a Multiattribute Training Data Mining Model for Basketball Players Based on Big Data**

Yunbin Li , Jinyan Ge, and Wei Hao




Research Article (14 pages), Article ID 6399266, Volume 2021 (2021)

**Integrated Design of Graduate Education Information System of Universities in Digital Campus Environment**

Jing Ma  and Bo Feng


Research Article (12 pages), Article ID 8357488, Volume 2021 (2021)

**An Improved MOEA/D Algorithm for Complex Data Analysis**

Weihua Qian , Jiahui Liu, Yuanguo Lin, Lvqing Yang , Jianwei Zhang , Hang Xu, Minghong Liao, Yuxuan Chen, Yunyi Chen, and Bingshuai Liu


Research Article (20 pages), Article ID 6393638, Volume 2021 (2021)

**Research on Performance Prediction of Technological Innovation Enterprises Based on Deep Learning**

Huan Liu 


Research Article (12 pages), Article ID 1682163, Volume 2021 (2021)

**Research on the Effect of Big Data Flipped Classroom Combined with Scenario Simulation Teaching: Based on Clinical Practice of Medical Students**

Shuli Liu, Yong Li, Xiaobo Wang, Xuena Zhang, and Renshu Wang 


Research Article (11 pages), Article ID 7107447, Volume 2021 (2021)

**Industry 4.0-Oriented Chipless RFID Backscatter Signal Variable Polarization Amplitude Deep Learning Coding**

Guolong Shi , Yigang He, Lichuan Gu, and Jun Jiao

Research Article (11 pages), Article ID 6985420, Volume 2021 (2021)

**5G Joint Artificial Intelligence Technology in the Innovation and Reform of University English Education**

Xia Sun 


Research Article (10 pages), Article ID 4892064, Volume 2021 (2021)



# Contents

---

**Research on Modeling and Scheduling Methods of an Intelligent Manufacturing System Based on Deep Learning**

Xiaoyi Lan and Hua Chen 


Research Article (11 pages), Article ID 4586518, Volume 2021 (2021)

**Artificial Intelligence System for College Students' Physical Fitness and Health Management Based on Physical Measurement Big Data**

Li Ai 

Research Article (10 pages), Article ID 4727340, Volume 2021 (2021)

**A Novel Stock Index Intelligent Prediction Algorithm Based on Attention-Guided Deep Neural Network**

Yangzi Zhao 


Research Article (12 pages), Article ID 6210627, Volume 2021 (2021)

**Neural Network Topology Construction and Classroom Interaction Benchmark Graph Based on Big Data Analysis**

Congcong Luan  and Peng Shang


Research Article (10 pages), Article ID 2334443, Volume 2021 (2021)

**College Oral English Teaching Reform Driven by Big Data and Deep Neural Network Technology**

Hui Liu 

Research Article (8 pages), Article ID 8389469, Volume 2021 (2021)

**Construction of Multimedia-Assisted English Teaching Mode in Big Data Network Environment**

Hongxin Zhao 


Research Article (10 pages), Article ID 1609187, Volume 2021 (2021)

**Stock Trend Prediction Algorithm Based on Deep Recurrent Neural Network**

Ruo Chen Lu  and Muchao Lu


Research Article (10 pages), Article ID 5694975, Volume 2021 (2021)

**Research on the Role of Big Data Technology in the Reform of English Teaching in Universities**

Xiaoge Jia 


Research Article (13 pages), Article ID 9510216, Volume 2021 (2021)

**Intelligent Learning Algorithm for English Flipped Classroom Based on Recurrent Neural Network**

Qi Shan 



Research Article (8 pages), Article ID 8020461, Volume 2021 (2021)

**Research on Flipped Classroom of Big Data Course Based on Graphic Design MOOC**

Yanqi Wang 

Research Article (11 pages), Article ID 4042459, Volume 2021 (2021)

**Feasibility of Using Improved Convolutional Neural Network to Classify BI-RADS 4 Breast Lesions: Compare Deep Learning Features of the Lesion Itself and the Minimum Bounding Cube of Lesion**

Meihong Sheng , Weixia Tang, Jiahuan Tang, Ming Zhang, Shenchu Gong, and Wei Xing 


Research Article (9 pages), Article ID 4430886, Volume 2021 (2021)

**Analysis of Japanese Expressions and Semantics Based on Link Sequence Classification**

Yanyan Shi  and Yuting Liang


Research Article (12 pages), Article ID 3389643, Volume 2021 (2021)

**Music Style Classification Algorithm Based on Music Feature Extraction and Deep Neural Network**

Kedong Zhang 


Research Article (7 pages), Article ID 9298654, Volume 2021 (2021)

**Principal Component Analysis and Prediction of Students' Physical Health Standard Test Results Based on Recurrent Convolution Neural Network**

Kai Hou 


Research Article (11 pages), Article ID 2438656, Volume 2021 (2021)

**Influencing Factors of Athletes' Injury Rehabilitation from the Perspective of Internal Environment**

Xiang Huang and Xiaoping Wang 


Research Article (7 pages), Article ID 2368847, Volume 2021 (2021)

**[Retracted] Construction and Analysis of Green Investment Risk Evaluation Index System Based on Information Entropy Fuzzy Hierarchical Analysis Model**

Yuqi Huang 


Research Article (13 pages), Article ID 4850321, Volume 2021 (2021)

**Optimization Study of Multidimensional Big Data Matrix Model in Enterprise Performance Evaluation System**

Honglin Fu 


Research Article (12 pages), Article ID 4351944, Volume 2021 (2021)

**Application of Flipped Classroom in the Era of Big Data: What Factors Influence the Effect of Teacher-Student Interaction in Oral English Teaching**

Yan Liu  and Wenjin Qi


Research Article (7 pages), Article ID 4966974, Volume 2021 (2021)

**Computer Vision-Driven Evaluation System for Assisted Decision-Making in Sports Training**

Lijin Zhu 

Research Article (7 pages), Article ID 1865538, Volume 2021 (2021)


**Application of Flipped Classroom Model Driven by Big Data and Neural Network in Oral English Teaching**

Yujun Zeng 

Research Article (7 pages), Article ID 5828129, Volume 2021 (2021)



# Contents

**[Retracted] Deep Convolutional Neural Network and Weighted Bayesian Model for Evaluation of College Foreign Language Multimedia Teaching**

Tingting Liu and Le Ning 




Research Article (7 pages), Article ID 1859065, Volume 2021 (2021)

**A Mental Health Assessment Model of College Students Using Intelligent Technology**

Keke Li  and Weifang Yu 



Research Article (10 pages), Article ID 7485796, Volume 2021 (2021)

**Research on a High-Speed and Heavy-Duty Closed-Loop Drive System of a Two-Phase Hybrid Stepping Motor Based on a Hybrid Controller**

Zhou Yansuo , Leng Yonggang , Lu Wenqi , Li Yu, Li Qingmian, and Wu Di


Research Article (13 pages), Article ID 2515820, Volume 2021 (2021)

**A Study on the Application of Interactive English-Teaching Mode under Complex Data Analysis**

Dongyang Xu  and Sang-Bing Tsai 


Research Article (12 pages), Article ID 2675786, Volume 2021 (2021)

**Flipped Classroom for Motor Skills: What Factors Influence College Students' Learning Effect?**

Fengyan Zhang, Baojuan Ma , and Wengang Ren


Research Article (7 pages), Article ID 2148905, Volume 2021 (2021)

**An Improved Genetic Algorithm and Neural Network-Based Evaluation Model of Classroom Teaching Quality in Colleges and Universities**

Huaying Zhang, Bin Xiao, Jinqiong Li, and Min Hou 

Research Article (7 pages), Article ID 2602385, Volume 2021 (2021)

**Convolutional Neural Network-Assisted Strategies for Improving Teaching Quality of College English Flipped Class**

Tiankun Liu 



Research Article (8 pages), Article ID 1929077, Volume 2021 (2021)

**Partial Color Photo Processing Method for Components Based on Image Enhancement Technology**

Hao Wu  and Zhi Zhou


Research Article (9 pages), Article ID 4132016, Volume 2021 (2021)

**Cloud Education Chain and Education Quality Evaluation Based on Hybrid Quantum Neural Network Algorithm**

Hong-Xia Liu , Yong-Heng Zhang, and Sang-Bing Tsai 


Research Article (11 pages), Article ID 1909345, Volume 2021 (2021)

**An Adaptive BP Neural Network Model for Teaching Quality Evaluation in Colleges and Universities**

Yong Jin , Yiwen Yang, Baican Yang, and Yunfu Zhang



Research Article (7 pages), Article ID 4936873, Volume 2021 (2021)

**Research on the Station Layout Method of Ground-Based Pseudolite Positioning System Based on NSGA-II Algorithm**

Li Yang, Kaiyuan Yang , and Danshi Sun






Research Article (7 pages), Article ID 1520859, Volume 2021 (2021)

**Sentiment Analysis of Chinese Paintings Based on Lightweight Convolutional Neural Network**

Jiaying Bian  and Xiaoying Shen 

Research Article (8 pages), Article ID 6097295, Volume 2021 (2021)

**Object Detection and Movement Tracking Using Tubelets and Faster RCNN Algorithm with Anchor Generation**

Prabu Mohandas , Jerline Sheebha Anni , Rajkumar Thanasekaran , Khairunnisa Hasikin , and Muhammad Mokhzaini Azizan 

Research Article (16 pages), Article ID 8665891, Volume 2021 (2021)

**Simulation of Sports Venue Based on Ant Colony Algorithm and Artificial Intelligence**

Rui Zhang , Weibo Sun , and Sang-Bing Tsai 





Research Article (11 pages), Article ID 5729881, Volume 2021 (2021)

**A College Student Behavior Analysis and Management Method Based on Machine Learning Technology**

Xiaoying Shen  and Chao Yuan 

Research Article (10 pages), Article ID 3126347, Volume 2021 (2021)

**A Heterogeneous Ensemble Learning Model Based on Data Distribution for Credit Card Fraud Detection**

Yalong Xie , Aiping Li , Liqun Gao , and Ziniu Liu 


Research Article (13 pages), Article ID 2531210, Volume 2021 (2021)

**Angle Estimation Using Local Searching for Bistatic MIMO Radar with Unknown MCM**

Chaochen Tang , Hongbing Qiu , Xin Liu , and Qinghua Tang 



Research Article (7 pages), Article ID 9938071, Volume 2021 (2021)

**Reader Scheduling for Tag Population Estimation in Multicategory and Multireader RFID Systems**

Zhiyong He 



Research Article (9 pages), Article ID 7901590, Volume 2021 (2021)

**Intelligent Recognition and Teaching of English Fuzzy Texts Based on Fuzzy Computing and Big Data**

Ling Liu  and Sang-Bing Tsai 

Research Article (10 pages), Article ID 1170622, Volume 2021 (2021)



**Online Data Migration Model and ID3 Algorithm in Sports Competition Action Data Mining Application**

Li Ju , Lei Huang, and Sang-Bing Tsai 


Research Article (11 pages), Article ID 7443676, Volume 2021 (2021)

# Contents



## **Distilling the Knowledge of Multiscale Densely Connected Deep Networks in Mechanical Intelligent Diagnosis**

Xiaochuan Wang , Aiguo Chen , Liang Zhang, Yi Gu , Mang Xu, and Haoyuan Yan  
Research Article (12 pages), Article ID 4319074, Volume 2021 (2021)





## **Evolutionary Algorithm for Multiobjective Optimization Based on Density Estimation Ranking**

Lin Li, Hengfei Wu , Xiujian Hu, and Guanglei Sheng  
Research Article (18 pages), Article ID 4296642, Volume 2021 (2021)






## **SOSPCNN: Structurally Optimized Stochastic Pooling Convolutional Neural Network for Tetralogy of Fallot Recognition**

Shui-Hua Wang, Kaihong Wu, Tianshu Chu, Steven L. Fernandes, Qinghua Zhou, Yu-Dong Zhang , and Jian Sun   
Research Article (17 pages), Article ID 5792975, Volume 2021 (2021)




## **Towards Effective Classification of aMCI Based on Resting-State Multiscale Brain Features and Machine Learning Approaches**

Chunting Cai , Jiqiang Yan, Yu Zhou, Wuyang Zheng , Chenhui Yang , Zhemin Zhang , Bokui Chen, and Dan Hong  
Research Article (10 pages), Article ID 9975237, Volume 2021 (2021)

## **Multiobjective Optimization regarding Vehicles and Power Grids**

Kaiyang Zhong , Ping Wang, Jiaming Pei , Jiyuan Xu , Zonglin Han , and Jiawen Xu   
Research Article (6 pages), Article ID 5552626, Volume 2021 (2021)








## **Recognition of Imbalanced Epileptic EEG Signals by a Graph-Based Extreme Learning Machine**

Jie Zhou , Xiongtao Zhang , and Zhibin Jiang   
Research Article (12 pages), Article ID 5871684, Volume 2021 (2021)


## **Evaluation and Prediction of COVID-19 Prevention and Control Strategy Based on the SEIR-AQ Infectious Disease Model**

Yue Yu , Yuxing Zhou, Xiangzhong Meng, Wenfei Li, Yang Xu, Manfeng Hu, and Jingxiang Zhang   
Research Article (12 pages), Article ID 1981388, Volume 2021 (2021)


## **Power Density Case Study for 5G mmWave Array Antennas**

Dianyuan Qi , Fangzhu Zou , Jing Zhao , Shaobin Sun , Huanbin Wei , Yiling Chen , and Zhan Xia   
Research Article (14 pages), Article ID 5512957, Volume 2021 (2021)

## **Association Analysis of Private Information in Distributed Social Networks Based on Big Data**

Dongning Jia , Bo Yin , and Xianqing Huang   
Research Article (12 pages), Article ID 1181129, Volume 2021 (2021)

## **A Novel Smart Depression Recognition Method Using Human-Computer Interaction System**



Lijun Xu , Jianjun Hou, and Jun Gao  
Research Article (8 pages), Article ID 5565967, Volume 2021 (2021)

### **A Violation Information Recognition Method of Live-Broadcasting Platform Based on Machine Learning Technology**

Xiaoying Shen  and Chao Yuan 

Research Article (10 pages), Article ID 2851483, Volume 2021 (2021)

### **ACEA: A Queueing Model-Based Elastic Scaling Algorithm for Container Cluster**

Kui Li , Yi-mu Ji , Shang-dong Liu, Hai-chang Yao, Hang Li, Shuai You, and Si-si Shao







Research Article (11 pages), Article ID 6621094, Volume 2021 (2021)

### **Social Network Big Data Hierarchical High-Quality Node Mining**

Dongning Jia , Bo Yin , and Xianqing Huang 


Research Article (10 pages), Article ID 1444755, Volume 2021 (2021)

### **Precision Measurement for Industry 4.0 Standards towards Solid Waste Classification through Enhanced Imaging Sensors and Deep Learning Model**

Leow Wei Qin, Muneer Ahmad , Ihsan Ali , Rafia Mumtaz , Syed Mohammad Hassan Zaidi , Sultan S. Alshamrani , Muhammad Ahsan Raza , and Muhammad Tahir

Research Article (10 pages), Article ID 9963999, Volume 2021 (2021)

### **A Denoising Autoencoder-Based Bearing Fault Diagnosis System for Time-Domain Vibration Signals**

Yi Gu , Jiawei Cao, Xin Song, and Jian Yao


Research Article (7 pages), Article ID 9790053, Volume 2021 (2021)

### **Chinese Personal Name Disambiguation Based on Clustering**

Chao Fan  and Yu Li



Research Article (7 pages), Article ID 3790176, Volume 2021 (2021)

### **An Efficient Q-Algorithm for RFID Tag Anticollision**

Lingyun Zhao, Lukun Wang , and Shan Du


Research Article (7 pages), Article ID 9967739, Volume 2021 (2021)

### **User Value Identification Based on Improved RFM Model and $K$ -Means++ Algorithm for Complex Data Analysis**

Jun Wu, Li Shi, Liping Yang, Xiaxia Niu , Yuanyuan Li, Xiaodong Cui, Sang-Bing Tsai , and Yunbo Zhang





Research Article (8 pages), Article ID 9982484, Volume 2021 (2021)

### **Entropy-Based Multiview Data Clustering Analysis in the Era of Industry 4.0**

Yi Gu  and Kang Li

Research Article (8 pages), Article ID 9963133, Volume 2021 (2021)

### **PPANet: Point-Wise Pyramid Attention Network for Semantic Segmentation**

Mohammed A. M. Elhassan , YuXuan Chen, Yunyi Chen, Chenxi Huang , Jane Yang, Xingcong Yao, Chenhui Yang , and Yinuo Cheng 

Research Article (16 pages), Article ID 5563875, Volume 2021 (2021)


# Contents

## **A Transfer Deep Generative Adversarial Network Model to Synthetic Brain CT Generation from MR Images**

Yi Gu  and Qiankun Zheng 



Research Article (10 pages), Article ID 9979606, Volume 2021 (2021)

## **Network Intrusion Detection Based on an Improved Long-Short-Term Memory Model in Combination with Multiple Spatiotemporal Structures**

Xiaolong Huang 


Research Article (10 pages), Article ID 6623554, Volume 2021 (2021)

## **Video Stream Session Migration Method Using Deep Reinforcement Learning in Cloud Computing Environment**

Lingling Li  and Huixia Liu 






Research Article (10 pages), Article ID 5579637, Volume 2021 (2021)

## **Multiobjective Optimization Method of Coevolution to Intelligent Agricultural Dynamic Services under the Internet of Things Environment**

Haihong Liang 

Research Article (10 pages), Article ID 5585062, Volume 2021 (2021)

## **CPEH: A Clustering Protocol for the Energy Harvesting Wireless Sensor Networks**

Yu Han , Jian Su , Guangjun Wen , Yiran He , and Jian Li 


Research Article (14 pages), Article ID 5533374, Volume 2021 (2021)

## **Coword and Cluster Analysis for the Romance of the Three Kingdoms**

Chao Fan  and Yu Li

Research Article (8 pages), Article ID 5553635, Volume 2021 (2021)

## **Microgrid Group Control Method Based on Deep Learning under Cloud Edge Collaboration**

Yazhe Mao, Baina He , Deshun Wang, Renzhuo Jiang, Yuyang Zhou, Xingmin He, Jingru Zhang, and Yanchen Dong

Research Article (8 pages), Article ID 6635638, Volume 2021 (2021)

## **An Empirical Study on Optimal the Allocations in Advertising and Operation Innovation on Supply Chain Alliance for Complex Data Analysis**

Jiang-Tao Wang , Jian-Jun Yu, Yu-Hsi Yuan , Sang-Bing Tsai , and Shu-Fen Zhang 


Research Article (11 pages), Article ID 6680300, Volume 2021 (2021)

## **Markdown Time for Perishables Based on Dynamic Quality Evaluation for Complex Data Analysis**

Jiang-Tao Wang, Jian-Jun Yu, Yu-Hsi Yuan , Sang-Bing Tsai , and Shu-Fen Zhang 

Research Article (8 pages), Article ID 6695626, Volume 2021 (2021)

## **A Fast Hybrid Strategy-Based RFID Tag Identification Protocol**

Xinyan Wang 

Research Article (7 pages), Article ID 6646812, Volume 2021 (2021)

## Retraction

# Retracted: Construction and Analysis of Green Investment Risk Evaluation Index System Based on Information Entropy Fuzzy Hierarchical Analysis Model

### Wireless Communications and Mobile Computing

Received 8 August 2023; Accepted 8 August 2023; Published 9 August 2023

Copyright © 2023 Wireless Communications and Mobile Computing. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their

agreement or disagreement to this retraction. We have kept a record of any response received.

### References

- [1] Y. Huang, "Construction and Analysis of Green Investment Risk Evaluation Index System Based on Information Entropy Fuzzy Hierarchical Analysis Model," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 4850321, 13 pages, 2021.



## *Retraction*

# **Retracted: Influencing Factors of Microlecture on the Teaching Effect of Ideological and Political Courses in Colleges**

### **Wireless Communications and Mobile Computing**

Received 8 August 2023; Accepted 8 August 2023; Published 9 August 2023

Copyright © 2023 Wireless Communications and Mobile Computing. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

### **References**

- [1] Y. Jin, "Influencing Factors of Microlecture on the Teaching Effect of Ideological and Political Courses in Colleges," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 6309221, 7 pages, 2021.

## Retraction

# Retracted: Deep Convolutional Neural Network and Weighted Bayesian Model for Evaluation of College Foreign Language Multimedia Teaching

### Wireless Communications and Mobile Computing

Received 1 August 2023; Accepted 1 August 2023; Published 2 August 2023

Copyright © 2023 Wireless Communications and Mobile Computing. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their





agreement or disagreement to this retraction. We have kept a record of any response received.

### References

- [1] T. Liu and L. Ning, "Deep Convolutional Neural Network and Weighted Bayesian Model for Evaluation of College Foreign Language Multimedia Teaching," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 1859065, 7 pages, 2021.

## Research Article

# A Framework for Identification and Classification of IoT Devices for Security Analysis in Heterogeneous Network

Hafiz Muhammad Zahid,<sup>1</sup> Yasir Saleem,<sup>1</sup> Faisal Hayat,<sup>1</sup> Farrukh Zeeshan Khan ,<sup>2</sup> Roobaea Alroobaea ,<sup>3</sup> Fahad Almansour,<sup>4</sup> Muneer Ahmad ,<sup>5</sup> and Ihsan Ali <sup>6</sup>

<sup>1</sup>Department of Computer Science & Engineering, University of Engineering and Technology, Lahore, Pakistan

<sup>2</sup>Department of Computer Science, University of Engineering and Technology, Taxila, Pakistan

<sup>3</sup>Department of Computer Science, College of Computers and Information Technology, Taif University, P. O. Box 11099, Taif 21944, Saudi Arabia

<sup>4</sup>Department of Computer Science, College of Sciences and Arts in Rass, Qassim University, Buraydah 51452, Saudi Arabia

<sup>5</sup>School of Electrical Engineering and Computer Science (SEECS), National University of Sciences and Technology (NUST), Sector H-12, 44000 Islamabad, Pakistan

<sup>6</sup>Department of Computer System and Technology, Faculty of Computer Science and Information Technology, Universiti Malaya, 50603 Kuala Lumpur, Malaysia

Correspondence should be addressed to Ihsan Ali; [ihsanalichd@siswa.um.edu.my](mailto:ihsanalichd@siswa.um.edu.my)

Received 12 June 2021; Revised 22 December 2021; Accepted 19 March 2022; Published 26 April 2022

Academic Editor: Antonio Guerrieri

Copyright © 2022 Hafiz Muhammad Zahid et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Internet of Things (IoT) is a promising technology enabling physical devices like cameras, home appliances, and other devices to communicate and interoperate with each other. The next wave transforms our homes, society, enterprises, and cities with the massive presence of IoT devices. The devices in the Internet of Things (IoT) may exchange sensitive data, and an important issue for any organization is to get the data secured and protected. The preliminary requirement for this is a mechanism detecting and reporting anomalies automatically to some central controller. Therefore, this mechanism should be able to classify legit IoT devices from unauthorized ones. Malicious IoT devices, non-IoT devices, and other types of man-in-the-middle traffic sources must be quarantined for noncompliance. This helps formulate administrative policies and regulate/police traffic in the network for better QoS management. This work proposed a framework-based hierarchical deep neural network (HDNNs) to distinguish IoT devices from non-IoT devices using a feature set of IoT-specific traffic. A system has been designed based on HDNN that classifies IoT devices to their specific categories and identifies new entrants with reasonable accuracy. The results show that HDNN can distinguish IoT and non-IoT devices with higher accuracy and as well as classify IoT devices into the respective classes with the required accuracy.

## 1. Introduction

In the modern age, billions of devices such as home appliances, traffic lights, and lampposts are connected to the Internet, also known as the Internet of Things (IoT) [1]. These devices, called IoT devices, have several sensors that generate valuable data. Communication among them is done under different protocols such as Wi-Fi, Bluetooth, ZigBee, and Ethernet, which helps IoT devices increase their functionality using actuators [2, 3] sometimes.

IoT devices are usually part of a heterogeneous network providing valuable services to society. The communication between these devices also facilitates aggregating and processing data and reacting to the environment's changes automatically. These devices are sometimes involved in sharing enormous data, forming a network that shares data at extremely high rates throughout and in a continuous fashion [4].

According to the reports of the International Data Corporation (IDC), the number of IoT devices will reach 41 billion by 2020, according to the report of the International

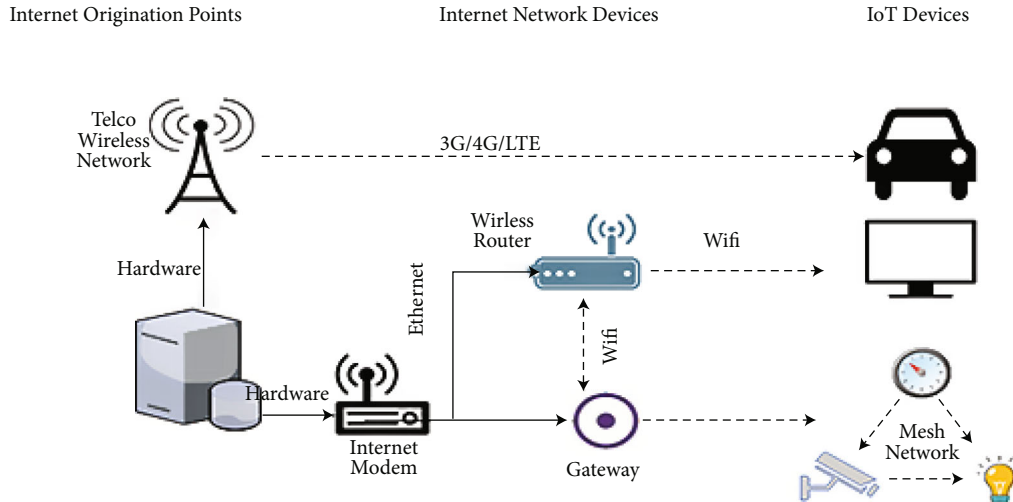


FIGURE 1: A heterogeneous network.

Data Corporation (IDC). The proliferation of IoT devices in IoT networks produces an operative challenge for administrators. The heterogeneous nature of IoT networks in big cities poses an additional asset management challenge as, in this case, the nature of IoT devices is diverse, and various departments install them. For instance, in such a heterogeneous network, light sensors can be fixed by the local council as shown in Figure 1. The local police division can install cameras, and different garbage and sewage sensors can be installed by the sanitation department. It is very difficult to identify the malfunctioning devices and their network location due to their invisibility [5].

The absence of direct human involvement differentiates the Internet of Things (IoT) from the traditional Internet. An IoT device can generate information using changes in the environment around, analyze it, and act upon it autonomously, however, with a price considering the data privacy, security, and protection [6]. Researchers have warned of the prospective risk of large numbers of unprotected devices communicating on the Internet. Therefore, developers and manufacturers have been struggling to develop a robust security system for IoT networks. In 2013, a researcher at the dev environment in organizational security service discovered the first IoT malware. According to the above analysis, more than 25% of the malware consisted of devices except for computers, such as smart cameras, smart TVs, smartwatches, and other home appliances [7].

Another problem is that the manufacturers of IoT devices do not provide regular updates for their devices unless users initiate firmware updates, owing to constrained resources [8]. These devices cannot run full-fledged security mechanisms. Therefore, IoT devices are prone to attacks (e.g., their default login passwords and unpatched bugs) for more extended periods [9].

IoT devices work mainly in an unattended environment, so there is a fair chance that an intruder may intentionally gain physical access to them. Resultantly, intruders may gain important information through a communication channel by secretly listening to the conversation because most IoT

devices use wireless links. These devices do not incorporate robust security features because computation and power resources are limited [10]. Implementation of solid security mechanisms is not possible due to the limited resources available and untrustful interaction with the environment. Considering the possibility of vulnerable IoT devices in an IoT network, there must be a robust security solution based on patching the vulnerabilities from time to time [11].

Nowadays, different organizations also facilitate IoT device connectivity, which might obtrude security threats to their networks. Organizations must be capable of determining the devices connected to their networks. They should provide a mechanism for identifying whether the connected devices in their networks are legit and do not pose a risk or threat [12].

Analysis of real-time network traffic has been used in several proposals for the identification of devices in general and for the classification of legit devices from nonlegit ones [13]. We can state that network traffic traces have been proven to differentiate IoT devices from non-IoT devices as there is a substantial difference in the data flow pattern of non-IoT devices compared to IoT devices. An IoT device may work when some trigger occurs. For example, object-detecting sensors work only when someone is passing in front of the sensor. However, considering IoT device classification alone, it is mostly very difficult to classify the network traffic of a device into a fixed pattern and to create an invariant profile even for the same types of IoT devices such as Drop Camera and Withings Smart Baby Monitor that are both cameras from different vendors or manufacturers. However, the traffic generated by these two cameras is another pattern, as shown in Figure 2.

In another scenario, as shown in Figure 2, the traffic generated from a Netatmo Weather Station is similar to the traffic generated by the baby monitor. Therefore, for a better classification of devices communicating in the heterogeneous network, it is important to identify a pattern that may help to place the devices in their respective category even if the devices generate the same kind of data.

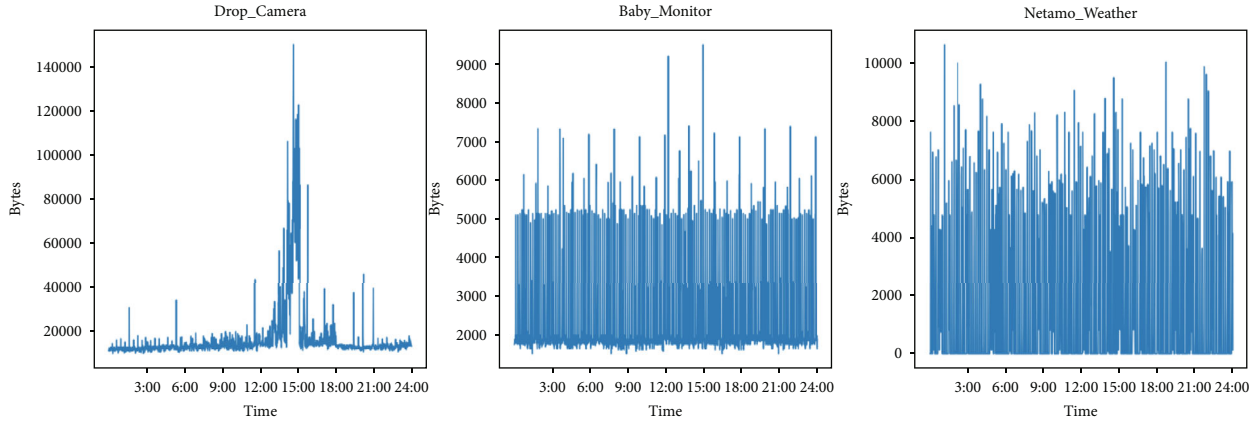


FIGURE 2: Daily traffic volume from three IoT devices [13].

IDE identification and classification of IoT devices in a heterogeneous network have been proposed based on port information, protocol, and MAC address [14]. By using the prefix (organization unique identifier (OUI)) of a MAC address, it is easy to identify the manufacturer of the device, but this information is not enough to guess the device functionality as many manufacturers produce diverse nature of devices. Additionally, manufacturers of different IoT devices may buy NICs from a third party, and it is impossible to know about IoT devices just by using the prefix (OUI) of the MAC address. Another aspect is this: The MAC address of IoT devices may even be spoofed [15].

In the recent past, multiple researchers have put forward many methods for automatic identification and classification of IoT devices with some constraints or rules [16]. These proposals mainly focused on the identification and classification of those IoT devices which have already been identified as part of the network under study. However, there is no proposal to identify and classify new devices entering the network. For example, when an unknown IP-based camera enters the network, it can be not easy to identify and classify it based on the previous IP base camera records.

Based on the discussion above, there is a need to build a holistic security solution for IoT networks. The preliminary requirement for this is the identification and classification of connected and incoming devices. In this paper, the challenge of identification and classification of IoT devices in the heterogeneous network has been taken up. We have used real-time network traffic and developed a method to differentiate IoT devices from non-IoT devices as well as classify IoT devices into their respective categories. IP addresses have not been used as a parameter. The reason is this: IP addresses can be spoofed very easily. We have used network traffic statistics and necessary metadata instead of relying on the deep packet inspection and proposed a hierarchical deep neural network (HDNN) framework that can discriminate between IoT devices from non-IoT devices and also identify the type of IoT devices with reasonable accuracy.

## 2. Literature Review

IoT is a network in which physical devices with sensors, software, and some other technologies act as tiny computers. These devices can collect data from the environment and process the data and then send this data to other devices with the help of internet connectivity. The exponential growth of IoT devices like smartwatches, smart cameras, lights, sensors, energy management devices, and other different types of devices in the intelligent environment has gained a ballistic interest of researchers to develop improved frameworks using ML and DL approaches to provide better services to human-kind in their lives. Researchers from different fields such as computer science, electrical, and other disciplines have paid great efforts to deal with this exponential growth. They are concerned about finding automated approaches for the security of IoT systems. In this section, we have covered closely related work done by different researchers on IoT for different categories, including IoT device identification, time-series data classification, and network traffic analysis using the manual as well as ML and DL approaches.

A lot of work has been done by researchers for characterizing internet traffic from 2005 to 2012. This work is primarily focused on web application detection such as Skype, mail, web browsing, and peer-to-peer applications but characterized IoT device traffic commonly known as mobile to mobile communication (M2M).

*2.1. Machine Learning Approaches for Traffic Analyses.* Much work has been done over the last few years using different types of machine learning and deep learning algorithms to classify traffic applications for specific computers and identify malware and botnets in the networks.

Lopez-Martin et al. [17] proposed a framework for a network traffic classifier (NTC) using a combination of DL models, CNN, and RNN. This framework was used to classify traffic flows like HTTP, SMTP, Telnet, YouTube, QUIC, and Office365 with six different features of traffic such as source port number, port number of destination, payload volume, transport control protocol (TCP) window size, interarrival time, and direction of traffic, and these features

were derived from first twenty packets of traffic flows. It was shown that the proposed framework with the combination of both CNN and RNN was better in the perspective of detection than other alternative algorithms without the need for any feature extraction or selection, which is mostly required when using different machine learning models.

In [18, 19], the authors used a combination of flow level features with different types of packet-level aspects such as packet size, byte payload distribution, packet interarrival times, and TLS handshake metadata (cipher suite code) to expand their work on the detection of malicious or illegal behavior on the network. Extraction of feature tools from the network was developed by researchers and was launched as an open-source tool for feature extraction.

*2.2. Machine Learning Approaches for IoT Identification and Classification.* Falk and Fries [18] proposed different types of authentication methods as a source of device identification and whitelisting (list of authorized devices). These methods were implemented for whitelisting in industrial automation control systems (IACS). Researchers found that in the IACS environment, the devices used in this domain were engaged in a communication relation that is already known. Therefore, the whole complexity of the system can be fixed. Authors noticed that large-scale enterprise environments are dynamic where new types of devices were frequently introduced. Thus, in this case, these methods can be failed.

Meidan et al. [19] applied the random forest (a machine learning approach) to extract features from network traffic data using feature extraction techniques as explained in [25] to identify an unauthorized device from many devices based on a single TCP flow in intelligent environments. Researchers collected data from 27 different IoT devices of nine different types and manually labelled the traffic data to train and evaluate a multiclass classifier for every device type. It was shown that it correctly discovered the unauthorized ninth device type and identified the remaining eight types of devices as a particular type on the list of authorized (white list) device types. This multiclass classifier used approximately 300 features (packet level and flow level). Among them, the essential attributes are lifetime minimum (TTL), median and average packets, the ratio of total bytes transferred and received, the total number of packets with reloading tag settings (RST), and Alexa server rank. This experiment's limitations were that researchers classified devices with specific device types, but there were many device types with a single device in its category. In this way, it cannot be generalized. The second drawback of this experiment was that the devices were identified with each other, but it was not for complex mixed real-time traffic.

Sivanathan et al. [5] proposed an approach for the classification of IoT and non-IoT devices using network traffic data collected over 3 weeks. The authors applied a random forest multiclass classifier to 12 attributes extracted from network traffic such as protocols, packet length, and port number and obtained a good accuracy for classification. This method has the drawback that it must be trained for each device using network traces, and this is not a practical approach for a large number of IoT devices in the commercial market.

Pêgo and Nunes [20] developed an application to discover the properties of a new device that can be used to decide the class of a device. This application automatically creates an interface and the required integration drivers for the new device. This paper's key concern was to identify the devices interacting within a network using the data exchanged by IoT devices. Researchers found the accuracy of different ML techniques for device discovery in the IoT smart environment. This forwarded a step towards automation of IoT devices in the IoT environment and reduced traditional device integration problems for platforms that bundle possible different IoT devices in an intelligent environment. The authors collected communication data by listening to smart environment traffic. This communication data (communication files with XML format) was converted to a database with information about each device in a smart environment using an application developed in the iPhone operating system (iOS) which applied different machine learning algorithms like the Levenshtein distance algorithm, TF-IDF tables, synonyms match, and finally and multi-property matching to discover the device that communicated in the IoT network correctly.

Ferrando and Stacey [21] described the issues and challenges to secure IoT devices. The authors proposed an approach for security detection applied to data streams and classified threats in the early stages. This approach is a step towards the novelty of securing IoT devices because this technique can classify the traffic generated by sensors and determine the diverse set of network anomalies. Researchers evaluated the method as anomaly detection based on data generated from a network device because most of the anomalies in network traffic data share-related attributes. The hypothesis was that noticing the distribution of features in network traffic was acceptable as examining the distributions of diagnostic power in the form of detection and classification of large categories of anomalies.

Shen et al. [22] explained how different supervised machine learning techniques could be applied to analyze data collected by listening to intelligent environment traffic and correctly identifying unauthorized IoT devices to protect the private information of an organization. Researchers trained and evaluated a multiclass classifier on the collected and manually labeled dataset from network traffic data of twenty-seven IoT devices of nine different types. They examined that it accurately identifies the ninth type as unknown and the remaining belonged to authorized devices.

Suárez and Salcedo [23] applied different classification techniques such as  $K$ -means and ID3 on the dataset collected from twelve different devices such as cameras, lights, sensors, and fridges. They used twelve features extracted from network communication data of IoT devices such as the capacity of the battery, size of memory, internet bandwidth required, gateway, Bluetooth enabled, etc. and determined four classes of devices using ML algorithms with the help of similar features of these devices.  $K$ -mean was tested on three, four, and five clusters and grouped the devices into four categories such as mobile orchestrators, fixed orchestrators, fixed followers, and dummy followers.

Lopez-Martin et al. [28] used deep learning approaches to classify the application layer protocols by using the features extracted from the data of packets captured at layer 3. Researchers tested the classification with many different sets of features, including both ports in some sets, win size, and payload size. In this paper, it has been shown the possibility of using the traffic rate to classify and identify information from network traffic.

Miettinen et al. [24] proposed a system capable of identifying the types of IoT devices automatically connected in the IoT smart environment. The authors used fingerprint classification that enabled enforcement of protocols and constraints to overcome damage as a result of unauthorized access to the network. The proposed system imposes some filtering traffic rules in the network to protect devices communicating in the smart environment due to threats originating from other highly risky devices in the network. The designed method was attempted by researchers to separate traffic for the IoT devices which were already seen in the network. This method is impractical because many IoT devices are being released every year. They have not used the method for mixed traffic generated from non-IoT devices. This method provides a way to generate and collect data.

Cvitić et al. [25] proposed a novel technique for the detection of distributed denial of services DDoS traffic generated by IoT devices, and this approach worked as a conceptual network model for anomaly detection. This model was based on the device classes and respective classes are totally dependent on the traffic generated by these devices separately.

In the last few years, different researchers worked on IoT device identification based on port information and MAC address. Nmap is an open-source tool that has robust functionality used to detect 2600 different versions of operating systems, but it is very difficult to guess the IoT device based on port information when IoT devices use HTTP or HTTPS ports as communication sources. Therefore, there must be a robust framework to identify and classify IoT devices and their categories based on traffic patterns generated from different devices in heterogeneous networks [26].

In this work, we have presented a framework that can discriminate between IoT devices from non-IoT devices and identify the type of IoT devices with the required accuracy according to a given traffic session or sequence of sessions.

### 3. Methodology

This research employs deep learning with a collection of different algorithms such as DBN, convolutional neural network, and DNN, inspired by the brain's functionality and structure. We have proposed a robust framework for IoT device identification and classification based on hierarchical deep neural networks using the Keras framework [27]. A type of artificial neural network with one input layer for input variables and one hidden layer and one output layer is known as a shallow neural network [28]. DNN is similar to the shallow neural network, but there is more than one hidden layer of neurons that process the inputs. All neurodes

of hidden layers are fully connected to all neurodes of input layers. Similarly, all neurodes of output layers are fully connected to all neurodes of previously hidden layers. Hidden layers are normally used for feature extraction or feature selection from features fed in the input layer. Neurodes in hidden layers act as feature detectors, and the number of hidden layers is increased; then hidden layers will be more optimal and more important features to the output layer for identification and Classification.

In ANN, there are three layers which are as follows:

- (i) Input layer (all inputs (features) provided to the model through this layer)
- (ii) Hidden layer (maybe more than one depending upon the problem and used for processing the inputs received from the input layer)
- (iii) Output layer (for prediction)

The input layer is used for communication with the external environment that provides the pattern to the neural network. This layer works with independent variables, and the neurons in this layer take decisions and fed them to the next hidden layer. The input layer must show the situation for which we have been training the neural network. Every neuron in the input layer represents independent variables with influence on the target variable in a neural network.

This hidden layer has a collection of neurons with different activation functions that can be applied to it, which can be found in between the input and output layers. It deals with the input layer's processed input, and its responsibility is just to extract the required features from the input data. There can be more than one hidden layer in DNN. The model's accuracy can be increased by increasing the number of neurons in the network and additional layers are useful up to a limit of 9-10. Accuracy may be constant or may be decreased as their predictive power can be declined. However, 3 to 10, mostly hidden layers, are being used nowadays. The number of neurons should be considered in each network, as the number of neurons depends on the problems' complexity. If there are unnecessary neurons in the network, then the model will lead to overfitting. If there are few neurons in the network, then these few neurons adequately detect the signal in the complex dataset.

The machine learning and deep learning models contain two types of parameters (hyperparameters and model parameters). Model parameters indicate how the input data can be used to get desired output by learning at training time whereas hyperparameters tell how our model can be defined at the start of training like how many hidden layers can be used in ANN. These hyperparameters can be decided as a judgment of an expert and can be changed concerning time for optimization, and similarly, model parameters like weights can be updated during backpropagation network for strong relationship or better accuracy.

The output layer receives the input from the hidden layer and executes it for identification and classification. It will check the predicted output with actual outputs, and if the difference between predicted and actual output is very

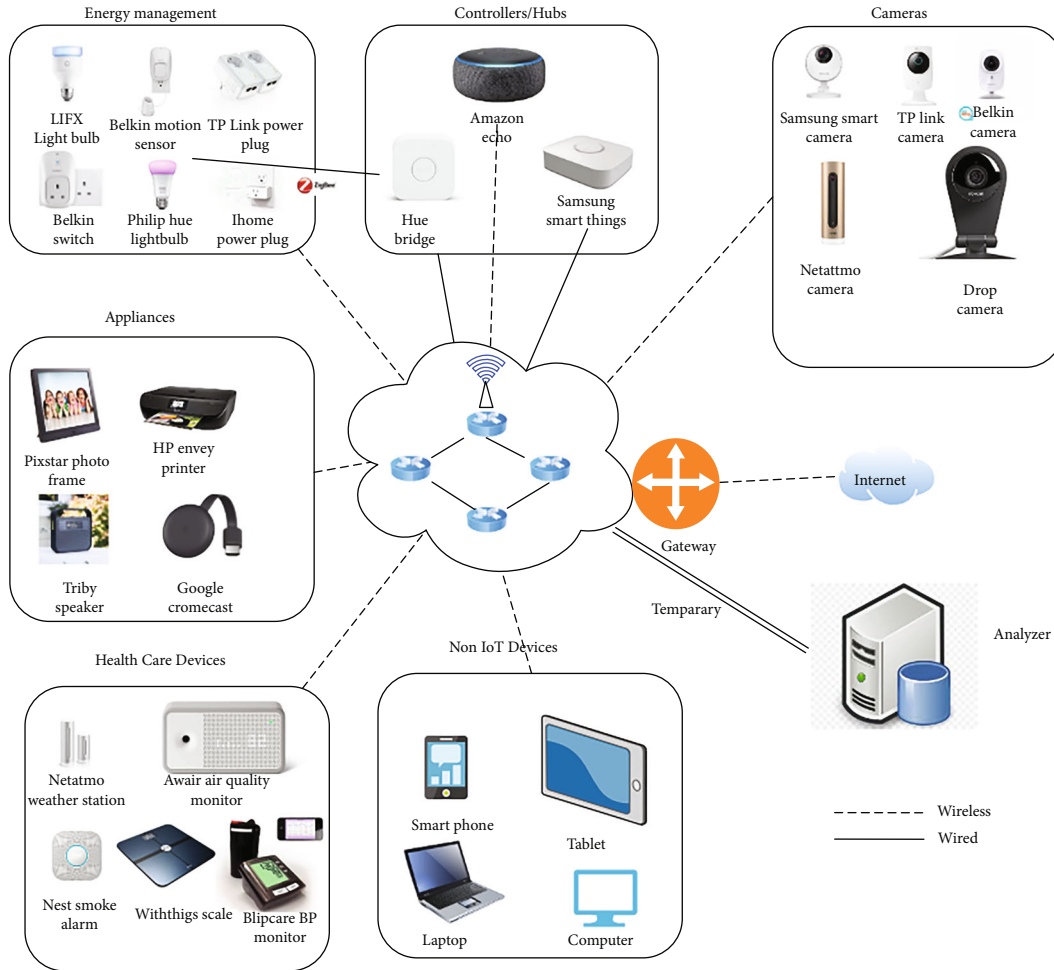


FIGURE 3: Testbed architecture for IoT and non-IoT devices.

high, then this layer traces the information back to the input layer and adjusts the weights by using a backpropagation network for every epoch. The number of neurons in the output layer must be according to the problem that is to be processed.

In this part of the paper, we have covered how pcap files (Wireshark files) are converted into executable (CSV) format. The dataset has been collected at the request of the University of New South Wales, Sydney (UNWS), Australia, as self-generated data in 28 pcap files collected from the testbed as a smart environment that has a number of IoT devices and non-IoT devices. The same dataset was employed by Bai et al. [13]. In the testbed, they used 28 IoT and non-IoT devices for communication, and their data were recorded under a synthesized network traffic trace for 6 months period.

Figure 3 shows the devices IoT and non-IoT used in the testbed.

As a first step, we converted the pcap files into an executable format (CSV format). For that purpose, the pcap files were transformed into CSV files using Python queries with 83 generalized features followed by data labeling for IoT device classification in a heterogeneous network. To label

our dataset with the specific device, we have used SQL server management studio. The dataset contains information about the network traffic stream of the devices that were used in the testbed for the collection of the dataset in a smart environment.

Using the large range of feature values in the data might lead to less accurate results and problems with the training. Hence, we have decided to use the built-in MinMaxScaler of the sklearn library in Python. This scaler can be used to perform min-max scaling, which will lead to the state that every value in the dataset is in the range (0, 1). We noticed that after performing the feature scaling, the test set results were better because the accuracy can be increased after the scaling of features.

Feature selection is one of the most important parts considered in IoT traffic for machine learning algorithms because this technique gives the most important and relevant features for target variables. Hence, the accuracy of the model can be increased. Feature selection also helps to reduce “the curse of dimensionality” that is well-known and might cause the model to overfit or perform poorly. Different machine learning algorithms do not often require feature selection like decision tree (DT) and random forest



TABLE 1: Optimal features after recursive feature elimination.

Features	Description
Source MAC	MAC address of the source
Source_port	Port number of source
Dest_MAC	MAC address of destination
Dest port	Destination port number
Flow ID	ID of network flow
Protocol	Protocols for communication (6 and 17)
Tot forw pack	Total forward packets
Flow duration	Total duration of network flow
Tot Back pack	Total number of packets backward
TotLen Bwd_Pkts	Length of all backward packets
TotLen Fwd_Pkts	Length of all forward packets
Forward_Pkt Len max	Max length of a forward packet from forward packets
Forward_Pkt Len min	Min length of a forward packet from forward packets
Forward_Pkt Len mean	Mean of forward packets
Forward_Pkt Len Std	STD of forward packets length
Backword_Pkt Len max	Greater length of a backward packet from all backward packets
Backward_Pkt Len_mean	Mean(average)length of backward packets
Backward_Pkt Len_Std	STD of backward packets length
Flow_Byts/s	Traffic flow in bytes/second
Flow_Pkts/s	Packets flow/second
Forward_Header Len	Length of header of forward packets
Backward_Header Len	Length of header of backward packets

(RF). The reason is that the feature selection process is being done on the fly due to the way these models are being trained (the “best” feature is selected at each split of the tree). However, some models may need feature selection to be performed to reach better results. In this work, we have used a hierarchical deep neural network, so we have to perform feature selection for that purpose. We have used different feature selection methods, but the recursive feature elimination (RFE) selected important and optimal features which we have used in our training dataset for the proposed framework. The features gave better results instead of other features.

Table 1 shows the most important and optimal features.

### 3.1. Experimental Evaluation

**3.1.1. Identification and Classification of Devices.** To identify and classify IoT devices for security analysis in heterogeneous networks, we have proposed an end-to-end robust framework based on hierarchical deep neural networks. Our proposed method is actually in two stages. In the first stage, we have used a deep neural network with one input layer, 4 hidden layers, and one output layer, all layers are fully connected, and it is used to distinguish between IoT devices from non-IoT devices such as laptops, MacBook, and Samsung Galaxy Tab. In the second stage, we have used a second deep neural network with 4 hidden layers with a different number of neurons, and it is used to classify IoT devices into their respective classes.

The mathematical model of the proposed deep neural network is given below [29]:

$$(x)^{(j)} = f \left( b + \sum_{i=1}^n x_i^{(j-1)} \times w_i \right), \quad (1)$$

$$P_t = \text{SoftMax} \left( x^{(j)} \right), \quad (2)$$

where  $b$  is bias,  $x$  is a vector that has input to the input layer,  $w$  is weights,  $n$  is the number of previous layer inputs,  $j$  is the number of hidden layers,  $P$  is a vector which has probability after SoftMax, and  $t$  represents the number of values in the  $P$  vector, if  $P_1 > P_2$ .

We find the total input for each neuron of the hidden layer and squash the total net input using the sigmoid activation function or logistic activation function. For the hidden layer, we have used the logistic activation function, and this same process will be repeated for all neurons of hidden layers and the output layer neurons. We can simplify Equation (1) as follows:

$$\text{net}_{h1} = \sum_{i=1}^n w_i \times x_i + b. \quad (3)$$

After the net input of each hidden layer, the neuron then squashes this net input using the logistic activation function to find the output of each neuron of the hidden layer, this

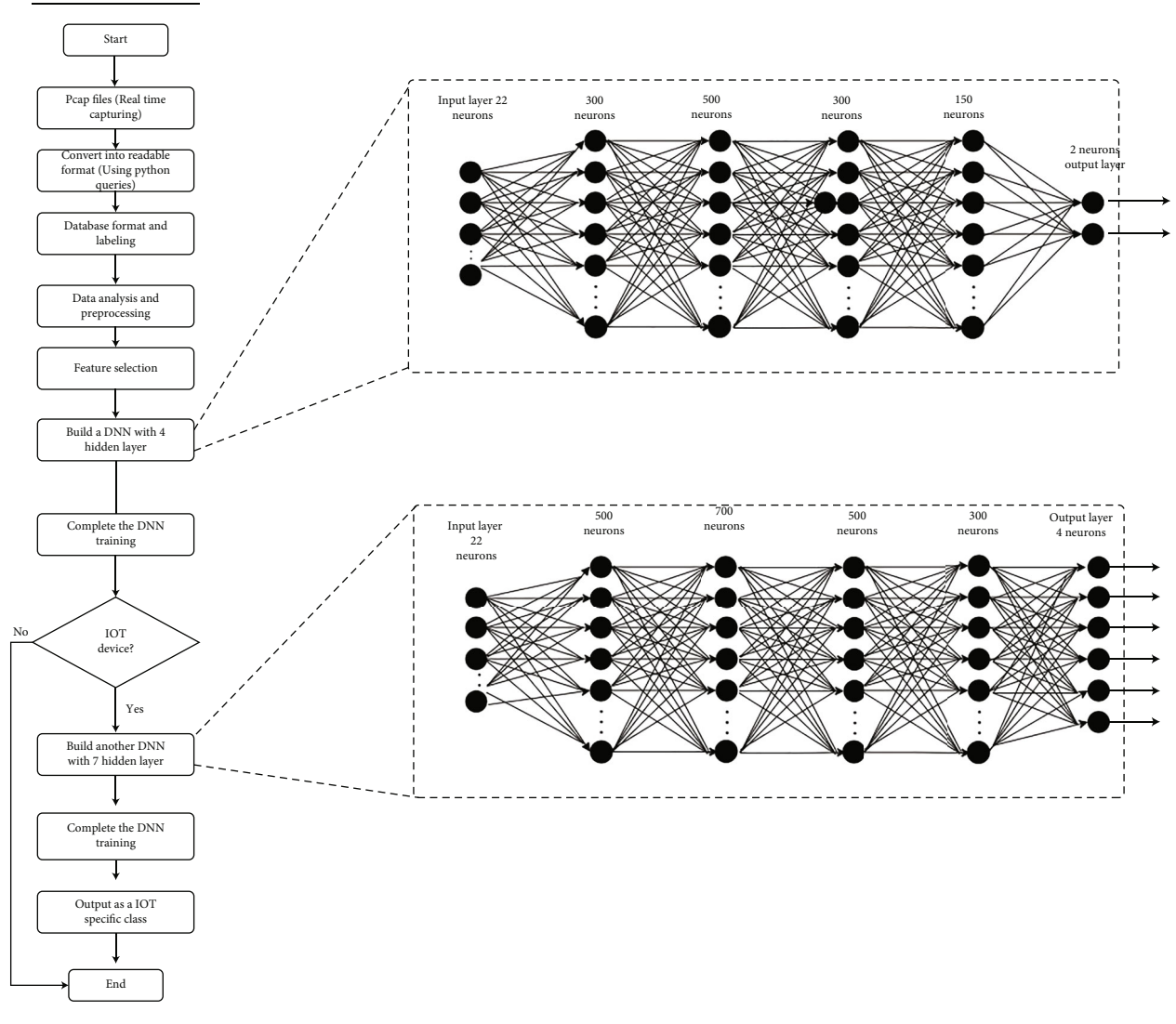


FIGURE 4: Proposed methodology with two DNNs.

same process can be used for all neurons of the output layer as well, and outputs of hidden layer neurons become the inputs of the output layer and at the output layer; we have used sigmoid function which can map value between 0 and 1.

$$out_{h1} = \frac{1}{1 + e^{-net_{h1}}}. \quad (4)$$

The error can be computed for each output neuron by using the squared error function and summed to get the total error:

$$E_{Total} = \sum \frac{1}{2} (target - actual)^2. \quad (5)$$

After getting the target value from the output of the output layer, we can calculate the error to adjust weights by using a backpropagation network.

Figure 4 shows the proposed methodology with two deep neural networks for IoT device classification.

The second deep neural network is also expressed as

$$(x)^{j1} = f \left( b1 + \sum_{i=1}^n x_i^{(j1-1)} \times w_i \right), \quad (6)$$

$$P_{t1} = SoftMax(x^{(j1)}),$$

where  $b1$  is bias,  $x$  is a vector that has input to the input layer,  $w$  is weights,  $n$  is the number of previous layer inputs,  $j$  is number of hidden layers,  $P$  is a vector which has probability after SoftMax, and  $t1$  represents the number of values in the  $P$  vector.

We find the total input for each neuron of hidden layer and squash the total net input using the sigmoid activation function or logistic activation function. For the hidden layer, we have used the logistic activation function, and this same

TABLE 2: List of devices with labelled data.

Category	Device name	Label	Connection type
Controllors and hubs	Samsung smart things	1	Wired connection
	Amazon Echo		Wireless connection
	Netatmo camera		Wireless connection
Cameras	Belkin camera	2	Wireless
	Samsung smart camera		Wireless
	Drop camera		Wireless
	TP-Link camera		Wireless
Switches and triggers	iHome power plug	3	Wireless
	Philip hue light bulb		Wireless
	Belkin switch		Wireless
	TP-link power plug		Wireless
Healthcare devices	Netatmo weather station	4	Wireless
	Awair air quality monitor		Wireless
	Nest smoke alarm		Wireless
	With things scale		Wireless
Electronics	Google Chromecast	4	Wireless
	HP envy printer		Wireless
Router	Tribby speaker	5	Wireless
	Bridge LAN (gateway) TP-Link router		6

TABLE 3: Nonoptimal hyperparameter list.

<i>First deep neural network</i>	
Input layer with tan $h$ activation function	22 neurons
Total hidden layers	3 hidden layers
First unseen (hidden) layer using the best activation function, Rectified Linear Unit (ReLU)	200 neurons
Second unseen (hidden) layer using the best activation function, Rectified Linear Unit (ReLU)	300 neurons
Third first unseen (hidden) layer using the best activation function, Rectified Linear Unit (ReLU)	100 neurons
Output layer with SoftMax activation function	2 neurons
Learning rate (LR)	0.01
Decay, momentum	$1e - 6, 0.9$
Loss, optimizer	mean_squared_error, sgd
Epochs	99
Batch_size	35
<i>Second deep neural network</i>	
Input layer with tan $h$ activation function	22 neurons
Total hidden layers	3 hidden layers
First unseen layer using a best activation function, Rectified Linear Unit (ReLU)	100 neurons
Second unseen (hidden) layer using a best activation function, Rectified Linear Unit (ReLU)	200 neurons
Third first unseen (hidden) layer using a best activation function, Rectified Linear Unit (ReLU)	100 neurons
Output layer with SoftMax activation function	6 neurons
Learning rate (LR)	0.01
Decay, momentum	$1e - 6, 0.9$
Loss, optimizer	categorical_crossentropy, sgd
Epochs	99
Batch_size	35

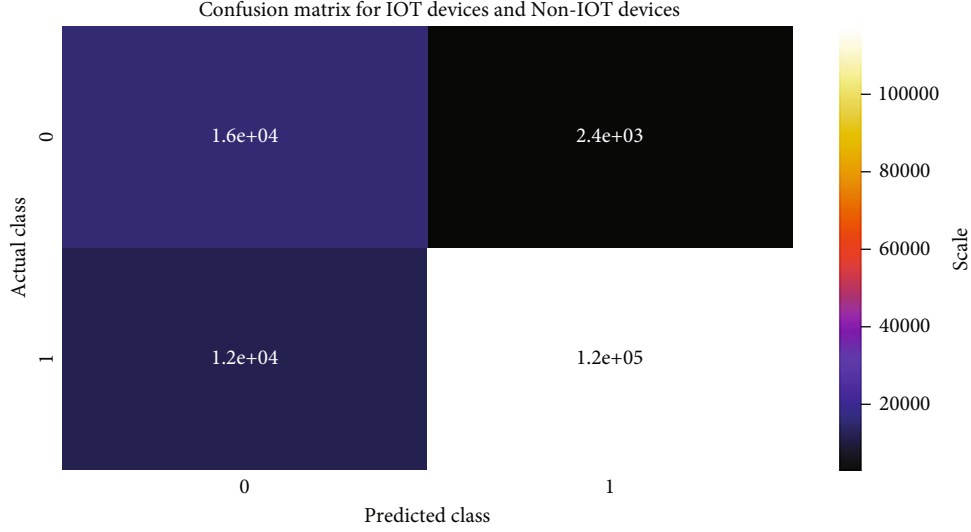


FIGURE 5: Confusion matrix of IoT vs. non-IoT devices.

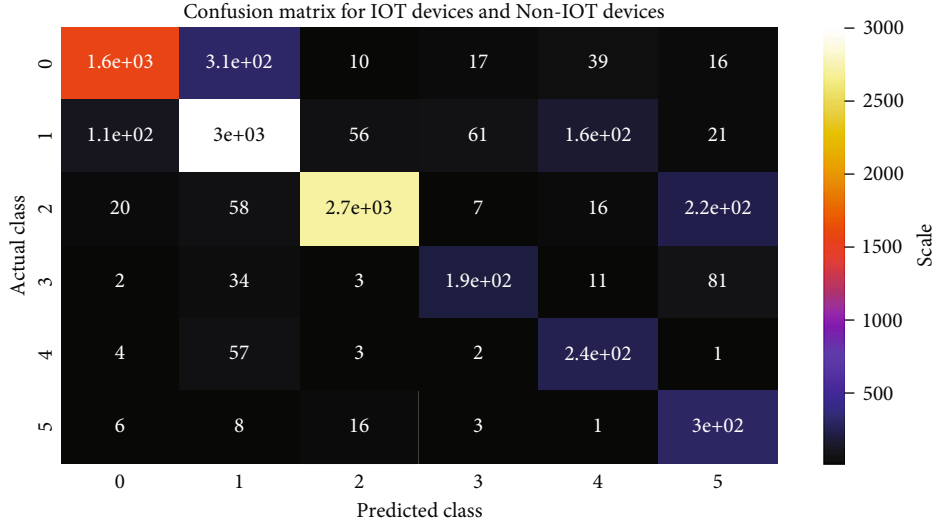


FIGURE 6: Confusion matrix for the 2nd phase of the framework.

process will be repeated for all neurons of hidden layers and the output layer neurons. We can simplify Equation (1) as follows:

$$net_{h1} = \sum_{i=1}^n w_i \times x_i + b. \quad (7)$$

After the net input of each hidden layer, neuron squashed this net input using the logistic activation function to find the output of each neuron of the hidden layer, this same process can be used for all neurons of the output layer as well, and the outputs of hidden layer neurons become the inputs of the output layer; and at the output layer, we have used sigmoid function which can map value between 0 and 1.

$$out_{h1} = \frac{1}{1 + e^{-net_{h1}}}. \quad (8)$$

The error can be computed for each output neuron by using the squared error function and summed to get the total error [30]:

$$E_{Total} = \sum \frac{1}{2} (target - actual)^2. \quad (9)$$

After getting the target value from the output of the output layer, we can calculate the error to adjust weights by using a backpropagation network.

Table 2 shows the devices with their specific categories and labels that we have used in our proposed method. This figure has not used the light bulb category in our proposed model because there is only one device in this category. Therefore, we have used only 6 categories for IoT devices labeled 1-6.

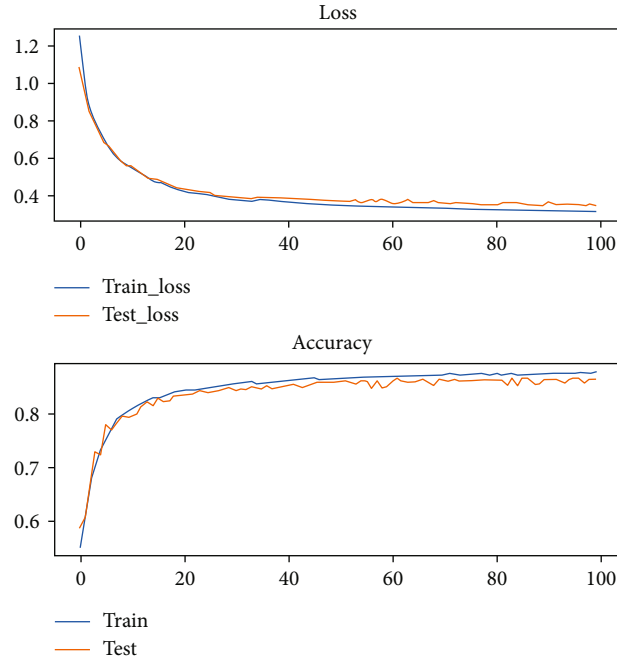


FIGURE 7: Accuracy and loss of the proposed framework for 99 epochs.

TABLE 4: Optimal hyperparameters list values.

<i>First deep neural network</i>	
Input layer with tan $h$ activation function	22 neurons
Total hidden layers	4 hidden layers
First unseen (hidden) layer using the best activation function, Rectified Linear Unit (ReLU)	300 neurons
Second unseen (hidden) layer using the best activation function, Rectified Linear Unit (ReLU)	500 neurons
Third first unseen (hidden) layer using the best activation function, Rectified Linear Unit (ReLU)	150 neurons
Fourth unseen (hidden) layer using the best activation function, Rectified Linear Unit (ReLU)	300 neurons
Output layer with SoftMax activation function	2 neurons
Learning rate (LR)	0.0001
Decay, momentum	$1e - 6$ , 0.9
Loss, optimizer	mean_squared_error, sgd
Epochs	3800
Batch_size	15
<i>Second deep neural network</i>	
Input layer with tan $h$ activation function	22 neurons
Total hidden layers	4 hidden layers
First unseen layer using the best activation function, Rectified Linear Unit (ReLU)	300 neurons
Second unseen (hidden) layer using the best activation function, Rectified Linear Unit (ReLU)	500 neurons
Third first unseen (hidden) layer using the best activation function, Rectified Linear Unit (ReLU)	700 neurons
Fourth unseen (hidden) layer using the best activation function, Rectified Linear Unit (ReLU)	300 neurons
Output layer with SoftMax activation function	6 neurons
Learning rate (LR)	0.00001
Decay, momentum	$1e - 6$ , 0.9
Loss, optimizer	categorical_crossentropy, sgd
Epochs	3800
Batch_size	30

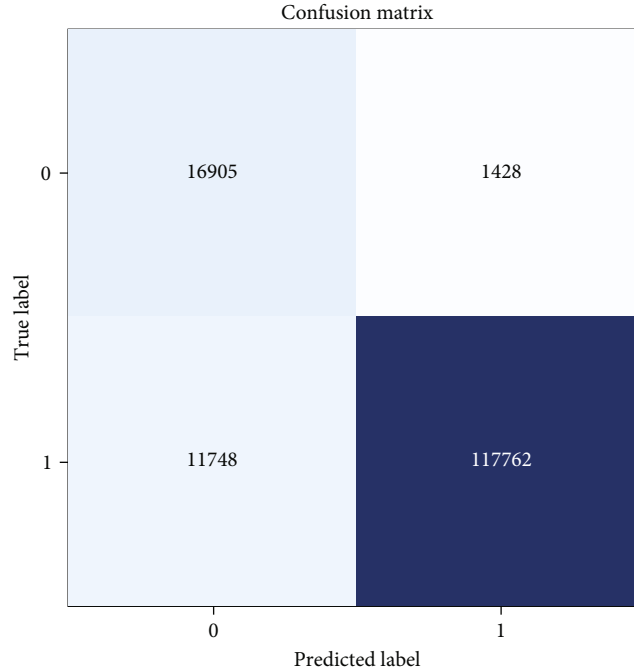


FIGURE 8: Confusion matrix for IoT and non-IoT devices.

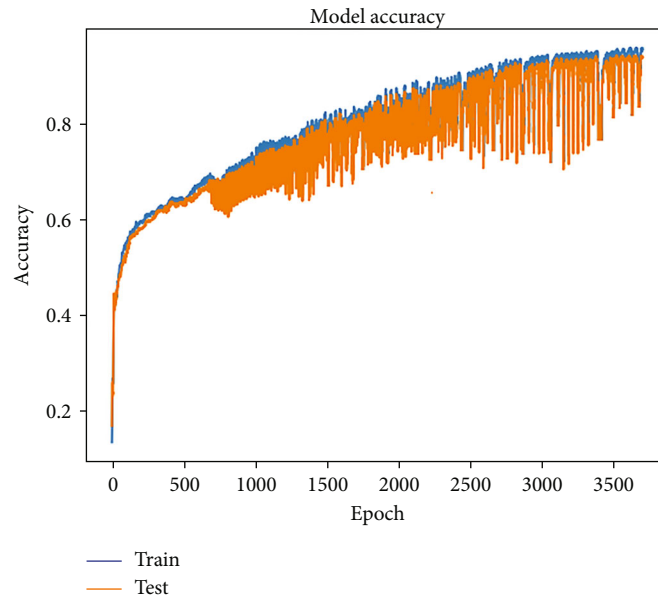


FIGURE 9: Training and testing accuracy of the first-stage DNN.

## 4. Performance Evaluation

4.1. *Data Construction and Modeling.* Data construction and modeling are as follows:

- (i) A total of 936893 samples were selected from the data pcap files for applying the hierarchical DNN technique
- (ii) 70% of training data for training the models and 30% of data for testing the model

(iii) We also used cross-validation on the training dataset with 10-folds

(iv) The hierarchical DNN technique is applied, and the respective accuracies are mentioned in the results

4.2. *Hyperparameter Setting for the Proposed Framework.* The models' parameters are used to describe a way of converting the input data into the model's desired output. Hyperparameters instead of model parameters are used to determine the structure of the model in use. The outcome

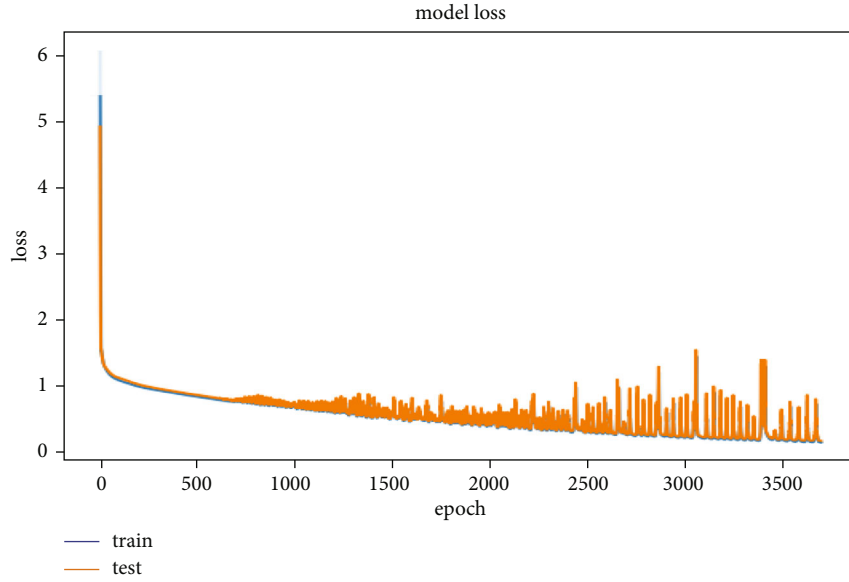


FIGURE 10: Loss of model with 0.2 of the first-stage DNN.

of the model can be changed by changing the values of hyperparameters. Therefore, the choice of hyperparameters is a crucial task and plays an important role. Keeping hyperparameters in mind is half part of the solution. The second part of the solution is knowing what kind of hyperparameter will be best for the model. In the proposed framework, we have set different hyperparameters according to requirements. Table 3 shows the values of nonoptimal hyperparameters that we have used in hierarchical deep neural networks but not getting high performance. Figures 5 and 6 represent the confusion matrix of the first- and second-stage hierarchical deep neural networks with nonoptimal hyperparameters. Figure 7 represents the accuracy and loss of the proposed framework for 99 epochs for nonoptimal hyperparameters.

Table 4 shows the optimal hyperparameters and achieved the required performance of the system.

In the proposed framework, we have applied the hierarchical DNN technique, and the respective performance measures are mentioned in the results.

Figure 8 describes the confusion matrix of the first-stage deep neural network with 4 hidden layers with a different number of neurons for optimal hyperparameters. In this figure, along the  $y$ -axis, the actual values are presented, and along the  $x$ -axis, predicted values are presented in which 0 shows non-IoT devices and 1 shows IoT devices.

The accuracy of the model can be measured as follows:

$$\begin{aligned} \text{Accuracy} &= \frac{\text{All right Diagonals}}{(\text{Total Number of samples})} \\ &= \frac{117762 + 16903}{147,841} \\ &= 0.9089(\text{with unseen data}), \end{aligned}$$

$$\text{Precision} = \frac{\text{True Positive}}{\text{Total predicted true}},$$

Confusion Matrix

0	1666	207	14	13	39	16
1	109	3014	42	61	159	21
2	20	58	2701	1	12	228
3	2	34	3	201	0	81
4	14	47	5	0	245	1
5	6	8	16	3	1	596
	0	1	2	3	4	5

FIGURE 11: Confusion matrix of the second-stage DNN for IoT device classification.

TABLE 5: Classification report.

Label	Support	Precision	Recall	F1-score
0	1817	0.85	0.92	0.88
1	3368	0.88	0.89	0.89
2	2781	0.89	0.97	0.93
3	279	0.63	0.72	0.67
4	456	0.79	0.54	0.64
5	943	0.95	0.63	0.76

$$\text{Recall} = \frac{\text{True Positive}}{\text{Total Actual true}}. \quad (10)$$

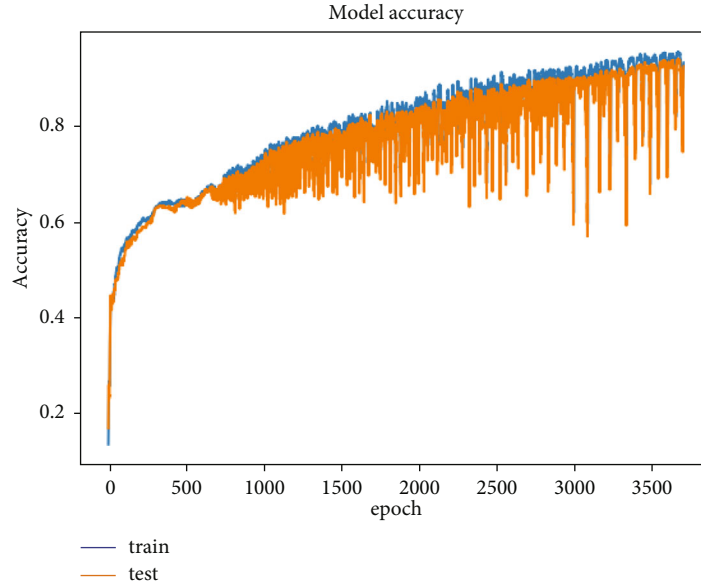


FIGURE 12: Training and testing accuracy of the second-stage DNN for IoT device classification.

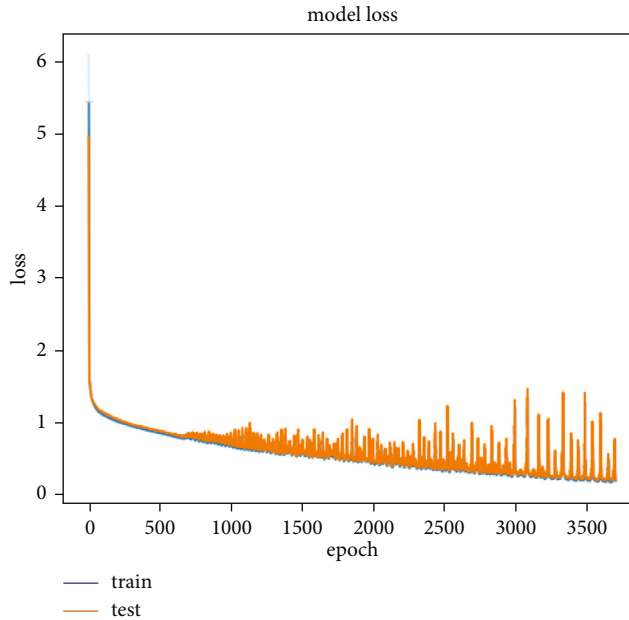


FIGURE 13: Training and testing loss of the second-stage DNN.

Figure 9 shows the accuracy (0.9120) curve for training as well as for testing the accuracy of the first-stage deep neural network (DNN) used to discernate IoT devices and non-IoT devices. Figure 10 shows the loss curve for training as well as for testing the loss of the first-stage deep neural network (DNN) used to distinguish between IoT devices from non-IoT devices.

Figure 11 shows the confusion matrix of the second-stage deep neural network used for IoT device classification using the heterogeneous network dataset.

Figure 11 describes the confusion matrix of the second-stage deep neural network with 4 hidden layers with a different number of neurons for optimal hyperparameters. In this

graph, along with the y-axis, the actual values are presented, and along the x-axis, the predicted values are presented within the range 0-5. Every value from 0 to 5 represents a particular class. The accuracy of the model can be measured as follows:

$$\begin{aligned}
 Accuracy &= \frac{(All\ Right\ Dignals)}{(Total\ Number\ of\ samples)} \\
 &= \frac{(1666 + 3014 + 2701 + 201 + 245 + 696)}{9288} \\
 &= 0.917958 \text{ (On unseen data),}
 \end{aligned}$$



TABLE 6: Comparison matrix.

Technique	Dataset	Result
Random forest (identification of devices)	UNSW	82.34%
Decision tree (identification of devices)	UNSW	79.88%
CNN-LSTM (categories [13])	UNSW	74.5%
Proposed framework (HDNN)	UNSW	91.30% (average)

$$Precision = \frac{(True\ Positive)}{(Total\ predicted\ true)},$$

$$Recall = \frac{(True\ Positive)}{(Total\ Actual\ true)}. \quad (11)$$

Table 5 shows a classification report of the model, which has different types of performance parameters. Precision can be calculated by dividing the true positive by the total number of predicted true as given in the equation. Similarly, recall can be calculated by dividing the true positives by the total number of actual trues as shown in the equation.

Figure 12 shows the accuracy curve for training and testing accuracy of the second-stage deep neural network (DNN) used to classify IoT devices with a validation accuracy of 0.9179. The  $y$ -axis represents the accuracy of the proposed model, and the  $x$ -axis represents the number of epochs. We trained our model for 3800 epochs. The blue curve shows the training accuracy, and the orange line shows the accuracy of the test dataset, which is clear that there is no overfitting, and the performance of the model is better. Figure 13 shows the loss curve for training as well as for testing the loss of the second-stage deep neural network (DNN) used to classify IoT devices into their respective class using traffic stream.

Table 6 shows a performance metric on existing techniques that we have applied and our proposed framework.

Previous systems were able to identify and classify IoT devices with a very small range of the dataset, and these were not end-to-end systems for a smart or intelligent environment. But the proposed HDNN is an end-to-end system that is used to discriminate IoT devices from non-IoT devices and, at the same time, classify IoT devices into their respective categories with reasonable accuracy, as shown in the proposed framework diagram. According to the comparison matrix, it is clear that the proposed system gave better performance than the previous ones.

## 5. Conclusion

With the emergence of the Internet of Things (IoT), a significant number of IoT devices are built in various areas, such as businesses, households, warehouses, and highways. The appropriate security of IoT devices is crucial since the state of different IoT devices has other properties. The literature survey demonstrated an excellent number of cited works on IoT device identification and classification. Still, most of the work applied to small enterprise networks is based on static information such as port information, MAC addresses, and model train test for devices.

The proposed framework based on hierarchical deep neural networks (DNNs) is used to discriminate IoT and non-IoT devices and classify IoT devices to their specific category. It has been shown that the proposed method is capable of an end-to-end system to distinguish IoT and non-IoT with 91% accuracy, besides classifying IoT devices to the respective classes with an accuracy of 91.33% in heterogeneous networks. According to the comparison matrix as shown in Table 6, it has been clear that already-proposed models were machine learning algorithms like random forest and decision tree for the classification of IoT devices in the environment in which only IoT devices were present.

However, proposed framework in this research can identify IoT devices and non-IoT devices and classify legitimate IoT devices into their specific classes with approximately 91% accuracy as an end-to-end system in a smart environment. This helps formulate administrative policies and regulate/police traffic in the network for better QoS management. In a future study, we aim to examine a broader range of IoT device types and non-IoT devices for building an intelligent environment, explore new communication technologies and as well as new deep learning techniques, and experiment with data from IoT devices compromised with spyware and cyber espionage and detection of unauthorized devices for security purpose.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

The authors are grateful to the Taif University Researchers Supporting Project (number TURSP-2020/36), Taif University, Taif, Saudi Arabia. This research work was also partially supported by the Faculty of Computer Science and Information Technology, University of Malaya, under Postgraduate Research Grant PG035-2016A.

## References

- [1] H. Tahaei, F. Afifi, A. Asemi, F. Zaki, and N. B. Anuar, "The rise of traffic classification in IoT networks: a survey," *Journal of Network and Computer Applications*, vol. 154, article 102538, 2020.

- [2] Z. Guan, J. Li, L. Wu, Y. Zhang, J. Wu, and X. Du, "Achieving efficient and secure data acquisition for cloud-supported internet of things in smart grid," *IEEE Internet of Things Journal*, vol. 4, no. 6, pp. 1934–1944, 2017.
- [3] D. Yu, L. Zhang, Y. Chen, Y. Ma, and J. Chen, "Large-scale IoT devices firmware identification based on weak password," *IEEE Access*, vol. 8, pp. 7981–7992, 2020.
- [4] M. Jindal, J. Gupta, and B. Bhushan, "Machine learning methods for IoT and their Future Applications," in *2019 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, pp. 430–434, Greater Noida, India, 2019.
- [5] A. Sivanathan, D. Sherratt, H. H. Gharakheili et al., "Characterizing and classifying IoT traffic in smart cities and campuses," in *2017 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPs)*, pp. 559–564, Atlanta, GA, USA, 2017.
- [6] Y. Yang, L. Wu, G. Yin, L. Li, and H. Zhao, "A survey on security and privacy issues in Internet-of-Things," *IEEE Internet of Things Journal*, vol. 4, no. 5, pp. 1250–1258, 2017.
- [7] B. Lam and C. Larose, *How did the internet of things allow the latest attack on the internet?*, Ed, 2016.
- [8] A. Alkhalil and R. A. Ramadan, "IoT data provenance implementation challenges," *Procedia Computer Science*, vol. 109, pp. 1134–1139, 2017.
- [9] R. Roman, J. Zhou, and J. Lopez, "On the features and challenges of security and privacy in distributed internet of things," *Computer Networks*, vol. 57, no. 10, pp. 2266–2279, 2013.
- [10] M. A. Al-Garadi, A. Mohamed, A. Al-Ali, X. Du, and M. Guizani, "A survey of machine and deep learning methods for Internet of Things (IoT) security," 2018, <http://arxiv.org/abs/1807.11023>.
- [11] F. Meneghello, M. Calore, D. Zucchetto, M. Polese, and A. Zanella, "IoT: Internet of threats? A survey of practical security vulnerabilities in real IoT devices," *IEEE Internet of Things Journal*, vol. 6, pp. 8182–8201, 2019.
- [12] M. Chiang and T. Zhang, "Fog and IoT: an overview of research opportunities," *IEEE Internet of Things Journal*, vol. 3, no. 6, pp. 854–864, 2016.
- [13] L. Bai, L. Yao, S. S. Kanhere, X. Wang, and Z. Yang, "Automatic device classification from network traffic streams of internet of things," in *2018 IEEE 43rd Conference on Local Computer Networks (LCN)*, Chicago, IL, USA, 2018.
- [14] P. Bajpai, A. K. Sood, and R. J. Enbody, "The art of mapping IoT devices in networks," *Network Security*, vol. 2018, no. 4, pp. 8–15, 2018.
- [15] A. Sivanathan, H. H. Gharakheili, F. Loi et al., "Classifying IoT devices in smart environments using network traffic characteristics," *IEEE Transactions on Mobile Computing*, vol. 18, no. 8, pp. 1745–1759, 2019.
- [16] A. Sivanathan, H. H. Gharakheili, and V. Sivaraman, "Managing IoT cyber-security using programmable telemetry and machine learning," *IEEE Transactions on Network and Service Management*, vol. 17, no. 1, pp. 60–74, 2020.
- [17] M. Lopez-Martin, B. Carro, A. Sanchez-Esguevillas, and J. Lloret, "Network traffic classifier with convolutional and recurrent neural networks for internet of things," *IEEE Access*, vol. 5, pp. 18042–18050, 2017.
- [18] R. Falk and S. Fries, "Using managed certificate whitelisting as a basis for internet of things security in industrial automation applications," *International Journal on Advances in Security*, vol. 8, no. 1 & 2, p. 2015, 2015.
- [19] Y. Meidan, M. Bohadana, A. Shabtai et al., "Detection of unauthorized IoT devices using machine learning techniques," 2017, <http://arxiv.org/abs/1709.04647>.
- [20] P. R. J. Pêgo and L. Nunes, "Automatic discovery and classifications of IoT devices," in *2017 12th Iberian Conference on Information Systems and Technologies (CISTI)*, Lisbon, Portugal, 2017.
- [21] R. Ferrando and P. Stacey, "Classification of device behaviour in internet of things infrastructures," in *Proceedings of the 1st International Conference on Internet of Things and Machine Learning*, New York, 2017.
- [22] J. Shen, Y. Li, B. Li, H. Chen, and J. Li, "IoT eye an efficient system for dynamic IoT devices auto-discovery on organization level," in *2017 IEEE 4th International Conference on Cyber Security and Cloud Computing (CSCloud)*, pp. 294–299, New York, NY, USA, 2017.
- [23] J. N. Suárez and A. Salcedo, "ID3 and k-means Based methodology for Internet of Things device classification," in *2017 International Conference on Mechatronics, Electronics and Automotive Engineering (ICMEAE)*, pp. 129–133, Cuernavaca, Mexico, 2017.
- [24] M. Miettinen, S. Marchal, I. Hafeez et al., "IoT sentinel demo: automated device-type identification for security enforcement in IoT," in *2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS)*, pp. 2511–2514, Atlanta, GA, USA, 2017.
- [25] I. Cvitić, D. Peraković, M. Periša, and M. Botica, "Novel approach for detection of IoT generated DDoS traffic," *Wireless Networks*, vol. 27, no. 3, pp. 1573–1586, 2021.
- [26] A. S. Hsu, *Automatic Internet of Things device category identification using traffic rates*, Virginia Tech, 2019.
- [27] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015, <http://arxiv.org/1511.06434>.
- [28] M. Lopez-Martin, B. Carro, and A. Sanchez-Esguevillas, "IoT type-of-traffic forecasting method based on gradient boosting neural networks," *Future Generation Computer Systems*, vol. 105, pp. 331–345, 2020.
- [29] O. Salman, I. H. Elhaji, A. Chehab, and A. Kayssi, "A machine learning based framework for IoT device identification and abnormal traffic detection," *Transactions on Emerging Telecommunications Technologies*, vol. 33, no. 3, article e3743, 2022.
- [30] I. Cvitić, D. Peraković, B. Gupta, and K.-K. R. Choo, "Boosting-based DDoS detection in Internet of Things systems," *IEEE Internet of Things Journal*, vol. 9, no. 3, pp. 2109–2123, 2022.

## Research Article

# Application of High-Resolution Remote Sensing Image for Individual Tree Identification of *Pinus sylvestris* and *Pinus tabulaeformis*

Hong Li<sup>1,2</sup> and Wunian Yang<sup>1</sup> 

<sup>1</sup>College of Earth Science, Chengdu University of Technology, Chengdu 610059, China

<sup>2</sup>Geology and Surveying Engineering School, Chongqing Vocational Institute of Engineering, Chongqing 402260, China

Correspondence should be addressed to Wunian Yang; [sherry676@139.com](mailto:sherry676@139.com)

Received 22 June 2021; Revised 24 July 2021; Accepted 4 August 2021; Published 19 October 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Hong Li and Wunian Yang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Aiming at applying unmanned aerial vehicle (UAV) remote sensing technology in extracting individual standing tree information, a new automatic single-tree information extraction method is proposed in this paper. The spectral information enhancement processing was performed on the original UAV image to highlight detailed local features; by importing DBI index, the optimal cluster number of the *K*-means clustering was automatically determined, and image pixels were then marked; Gauss Markov random field (GMRF) model was employed to segment the image further; by mathematical morphology, operators to postprocess the segmentation results to obtain the individual standing tree crown information, and individual standing tree position was calculated through image geometric moment as the basis for its identification. The results show that with the proposed extraction method, the overall accuracy of standing tree identification for the *Pinus sylvestris* and *Pinus tabulaeformis* forest areas are 95.65% and 89.52%; the single-tree crown extraction accuracy is 95.65% and 81.90%, respectively. This method exhibits good applicability while it does not require a large amount of manual intervention and prior knowledge, which significantly improves the automation of information extraction.

## 1. Introduction

The extraction of forestland information is one of the most important fields of remote sensing technology applications. The information extraction technology of individual standing trees mainly refers to the identification of individual standing tree and extraction of individual tree crowns, which provides a practical and effective scientific basis for investigating the forestland tree species [1], density estimation, planting survival rate [2, 3], change monitoring, ecological resource protection, etc. [4]. The traditional machine vision features include hog, sift, surf, orb, LBP, and Haar. These features can only extract limited information, not enough to support the subsequent detection task.

The standing tree information is traditionally obtained by manual field measurement, which is slow with high cost

[5]. The application of remote sensing technology and methods not only improves the efficiency of an individual [6, 7] standing tree information extraction but also provides a broader range of data sources including high spatial resolution satellite image data [7, 8], airborne hyperspectral, multispectral image data, and airborne radar data (such as IKONOS, QuickBird, and WorldView series) [9]. For some remote and barren research areas, it is often challenging to acquire corresponding high spatial resolution satellite remote sensing image in time, which limits its application in monitoring [10], and the method of airborne sensors is complicated and risky. In recent years, UAV (unmanned aerial vehicle) remote sensing technology has been in rapid development. UAV has the outstanding characteristics of convenient usage and maintenance, small equipment size, low cost, low loss, low risk, fast image acquisition speed, and high resolution [7, 11], which is

practically advantageous in regional remote sensing monitoring. In addition, the flying height of UAVs is generally low, so they are rarely affected by the cloud in the process of data acquisition, and the flight route can be flexibly planned and particularly designed [12–14]. Target detection technologies mainly include one-stage and two-stage technologies, including RCNN, fast RCNN, and fast RCNN in one stage, and Yolo and SSD in the other stage.

At present, scholars have carried out researches on the extraction methods for individual standing trees; most studies are based on different types of remote sensing data from various perspectives. From high spatial resolution satellite image data, the researchers used the extracted local maximum value to identify individual standing trees. The local maximum method usually takes the maximum value of the local spectrum as the centre of the tree crown and regards it as the position of the tree. This conventional method is relatively simple, and the extraction speed is fast; however, when the image brightness value changes much, or the image background is complex, it generates unsatisfactory recognition output [15]. Other studies report by combining object-oriented method and hydrological analysis technique [16], single-tree information extraction method is applied to high spatial resolution satellite image data. However, the traditional object-oriented methods often require excessive manual interventions [5, 14]; thus, it is difficult to fulfil the automatic extraction [17] of individual standing tree information. Spatial resolution refers to the size or size of the smallest unit that can be distinguished in detail on the remote sensing image, which is used to characterize the image to distinguish the details of ground targets. It is usually expressed by pixel size, resolution, or field of view.

In this paper, based on high spatial resolution UAV remote sensing image data, a new method for extracting standing tree information from a single plant was proposed; the individual standing tree information was extracted and analyzed, in comparison with other data sources and traditional theoretical methods. The accuracy of the method is verified and analyzed by selecting a specific research area.

## 2. Study Area and Data

**2.1. Overview of the Study Area.** The study area is located in the Shandong mining area in Daliuta Town, Shenmu County, Yulin City, Shaanxi Province. The satellite base map is the Gaofen-2 satellite image of the study area on July 08, 2020. The Shandong mining area is rich in mineral resources with sparse vegetation, fragile ecological environment, frequent sandstorms, droughts, and severe wind erosion.

Since its development and construction in the 1980s, its ecological restoration has been implemented through the combination of water and soil conservation and biological measures. Artificial afforestation is mainly conducted by planting evergreen trees such as *Pinus sylvestris* and *Pinus tabulaeformis* and shrubs such as sea buckthorn. In this paper, two representative areas are selected as experimental areas, as shown in Figure 1. The geographical range of experimental area 1 (with *Pinus sylvestris*) is  $110^{\circ} 12' 13''$  E– $110^{\circ} 12' 21''$  E,  $39^{\circ} 18' 36''$  N– $39^{\circ} 18' 42''$  N,  $0.03 \text{ km}^2$ , an aver-

age crown width of 1.0 m; the geographic area of experimental area 2 (with *Pinus tabulaeformis*) is  $110^{\circ} 13' 09''$  E– $110^{\circ} 13' 20''$  E,  $39^{\circ} 15' 31''$  N– $39^{\circ} 15' 41''$  N,  $0.08 \text{ km}^2$ , and an average crown width of 1.5 m.

**2.2. Data Acquisition and Processing.** This research uses a DJI Phantom 4 PRO drone, which is a consumer-grade quadrotor drone product. This mode is battery powered with a hovering shooting function and can fly for about 30 mins. The UAV platform is equipped with a CMOS (complementary metal oxide semiconductor) digital camera with a field view  $84^{\circ}$ ; the recording resolution is  $1920 \times 1080$  pixels with R (red light), G (green light), B (blue light) spectral channels.

Two field drone aerial photography experiments were carried out in August 2020 and September 2020, respectively. During the experiment, the weather was clear, and the wind was low. The data is collected by video shooting; the flying height of the drone during aerial photography is about 50 m.

By the MATLAB R2017a software platform, a video frame is extracted from UAV aerial video, and the TIFF format image file is saved as the image data for research. The data only contains three spectral channels of R, G, and B [18]; the spatial image resolutions of *Pinus tabulaeformis* area and *Pinus tabulaeformis* area are about 0.05 m and 0.03 m. The typical regional aerial image obtained is shown in Figure 2, which is further used as the image data for the extraction of tree information in this paper. TIFF (tag image file format) image file is one of the commonly used formats in graphics and image processing. Its image format is very complex, but because it stores image information flexibly, it can support many colour systems, and it is independent of the operating system, so it has been widely used. In various applications such as GIS, photogrammetry, and remote sensing, the image is required to have geocoding information, such as the coordinate system, scale, coordinates of points on the image, longitude and latitude, length unit, and angle unit.

## 3. Research Methods

**3.1. Individual Tree Information Extraction.** The flow chart of the information extraction of individual standing trees based on the high spatial resolution UAV remote sensing image is shown in Figure 3, the left columns display technical methods used, and the operational targets to be achieved are listed in the middle.

**3.1.1. Image Spectral Information Enhancement.** The extraction of a single tree based on remote sensing images can essentially be regarded as the process of separating foreground and background on images [19], that is, image segmentation. In this study, the foreground refers to the area covered by a single standing tree, while the background refers to the soil, grassland, and other areas. The image enhancement technology is considered conducive to the separation of the foreground and background in the image [18–20]. The method of image enhancement is to add some

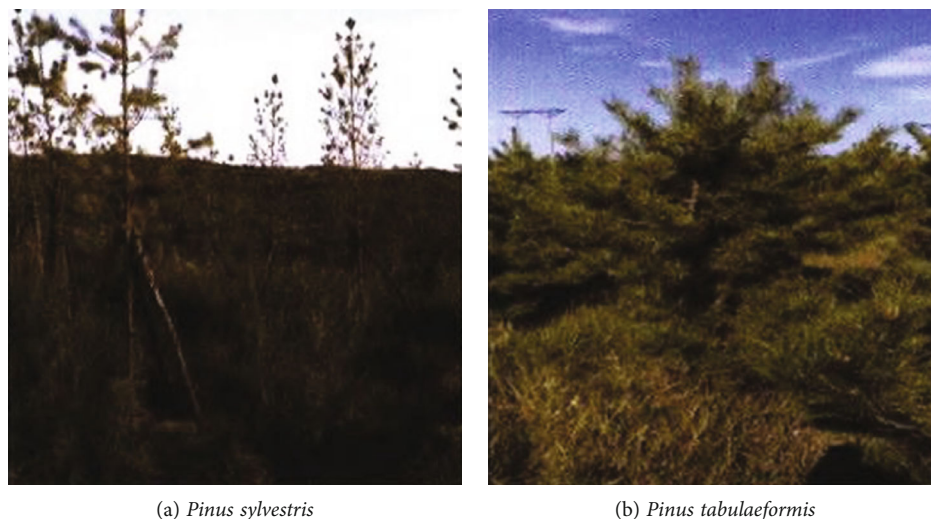


FIGURE 1: Pictures of experimental areas 1 and 2.

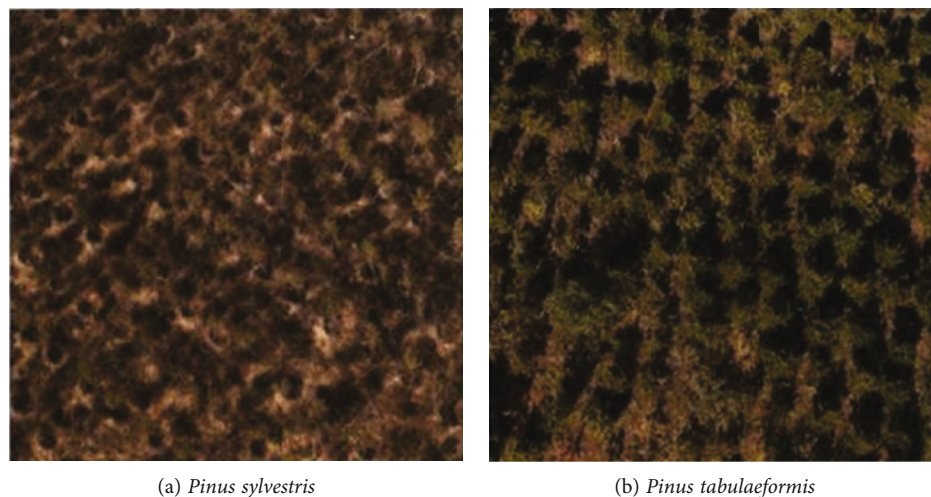


FIGURE 2: Aerial images of areas 1 and 2.

information or transform data to the original image by certain means, selectively highlight the features of interest in the image, or suppress (mask) some unnecessary features in the image, so as to make the image match the visual response characteristics. In the process of image enhancement, the reason of image degradation is not analyzed, and the processed image is not necessarily close to the original image. Image enhancement technology can be divided into two categories: spatial domain-based algorithm and frequency domain-based algorithm.

Since this study uses true colour image data, there exists a high degree of correlation between various bands [21]. When the traditional contrast stretching method is used to enhance the image, it can only improve the image chroma and brightness, but cannot perfect the image colour saturation degree. However, the decorrelation stretching method with a principal of component transformation can enlarge the coupling degree of the image information within the

high band, thereby increasing the image colour saturation. In addition, the colours of the relevant areas of the image are more vivid and prominent, while the dark areas will become brighter. Therefore, this paper adopts the decorrelation stretching spectrum enhancement method to operate the original images.

**3.1.2. Image Pixel Mark.** For the purpose of separation for image foreground and background, further image pixel labelling is required on the basis of image enhancement processing. During this process, the  $K$ -means clustering is used in the image pixel labelling process. The purpose of image pixel clustering is to group or classify the pixels in the images. The clustering algorithm is to divide a data set into multiple classes or clusters by a certain standard, such as distance, so that the data objects in the same cluster are similar, and the data objects in different clusters are different as much as possible. In addition, the commonly used clustering

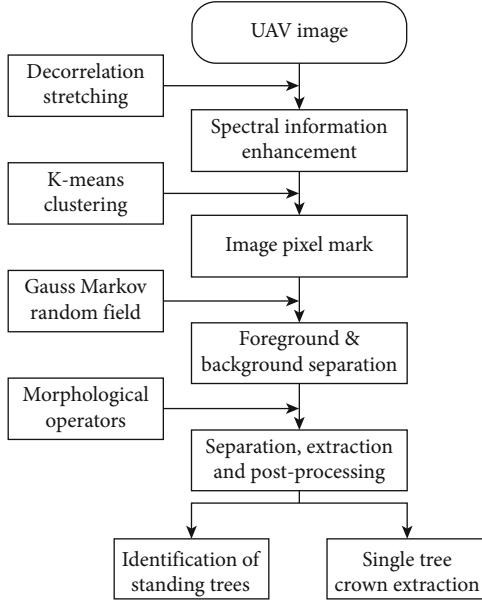


FIGURE 3: The flow chart of the information extraction.

algorithms include  $K$ -means,  $K$ -modes,  $K$ -prototypes, dmscan density clustering, GMM, and hierarchical clustering.

Assuming the image is a spatial data set  $X$ , the elements  $x_i$  ( $i = 1, 2, 3, \dots, N$ ) in the data set correspond to the pixels in the image [22]. In the theory of  $K$ -means clustering,  $K$  initial clustering centres  $\mu_k$  ( $k = 1, 2, 3, \dots, K$ ) are to be found in the sum of squares of Euclidean distances between clustering centres and image pixels reaches the minimum, and the pixels are assigned to the corresponding cluster centres. The objective function is as follows:

$$D = \sum_{i=1}^N \sum_{k=1}^K \|x_i - \mu_k\|. \quad (1)$$

For the conventional  $K$ -means clustering method, it is necessary to artificially set the initial clustering centre several times to obtain the optimal result, which reduces efficiency and is affected by human subjectivity to a certain extent. Therefore, the introduction of the DBI index in this study can automatically determine the optimal number of clusters, avoid subjective judgments, and improve clustering accuracy. DBI mathematically calculate the ratio of the intraclass distance and the interclass distance between any two categories [23]. The larger the DBI value is, the smaller the intraclass distance and the larger the interclass distance would be. The calculation formula is

$$d_{i,j} = \|v_i - v_j\|, \quad (2)$$

$$S_i = \frac{1}{|C_i|} \sum_{x \in C_i} \|x - v_i\|, \quad (3)$$

$$\text{DBI} = \frac{1}{K} \sum_{i,j \in [1,K]} \max \left( \frac{S_i + S_j}{d_{i,j}} \right). \quad (4)$$

In the formula,  $x$  represents data objects in the  $i_{\text{th}}$  class;  $v_i$  and  $v_j$  denote the centroids of the  $i$  and  $j$  classes, respectively;  $C_i$  is the number of data objects in the  $i_{\text{th}}$  class;  $d_{i,j}$  represents the Euclidean distance between the centroid of the  $i_{\text{th}}$  class and  $j_{\text{th}}$  class;  $S_i$  and  $S_j$  are the class distance between each data object in  $i_{\text{th}}$  and  $j_{\text{th}}$  class and centroids  $v_i$  and  $v_j$ , and  $K$  is the number of clusters.

Therefore, in this study, the process of image pixel labelling method is described as follows:

- (1) Preset  $K$  value ( $K \in [K_{\min}, K_{\max}]$ ), where  $K_{\min}$  is 2, and  $K_{\max}$  is 15, that is, the image clustering categories are divided into at least two categories, and at most 15 categories
- (2) Initialize the cluster centre  $C_k$ , ( $k = 1, 2, 3, \dots, K$ ) and calculate the  $D$  value, then redistribute the image pixels
- (3) Calculate the new cluster centre according to the formula below:

$$\mu'_k = \frac{1}{|C_k|} \sum_{x \in C_k} x \quad (5)$$

- (4) Repeat steps (2) and (3) until the cluster centre no longer changes. Record the corresponding DBI values, and label the corresponding category of image pixels, mark value of  $f \in [1, K]$
- (5) Produce DBI, when the number of clusters corresponds to the minimum value of DBI, the optimal number of clusters  $K_s$  is obtained, and output of corresponding image pixel marking result is given. The shortcoming of  $K$ -means are as follows: (1) The selection of  $K$  value is not easy to grasp. (2) It is difficult to converge for nonconvex data sets. (3) If the data of each hidden category is unbalanced, such as the data quantity of each hidden category is seriously unbalanced, or the variance of each hidden category is different, the clustering effect is not good. (4) Using iterative method, the result is only local optimal

**3.1.3. Image Foreground/Background Separation.** When using the  $K$ -means clustering method for image pixel labelling, only the grey value information of image pixels is considered, while the statistical dependence between neighbouring pixels is not taken into account, which renders some certain errors of labelling results. Hence, according to the Gaussian Markov random field (GMRF) theory [24], the spatial characteristics of image pixels were also considered, and image segmentation was carried out again on the marked image results. GMF is an image segmentation model built on the image neighborhood system; that is, through a neighborhood system, the pixels in the image and other pixels in its neighborhood are connected. The image to be segmented is described by constructing two

random fields, a marker field and a feature field [22]. It should be noted that the prior probability is the probability of current event based on historical experience, posterior probability is the probability of an event based on data or evidence, and likelihood probability is the probability of the occurrence of data or evidence when the probability of an event is known. The marker field is used to mark image pixels and to describe the result of image segmentation, denoted as  $Y = \{Y | y_1, y_2, \dots, y_n\}$ ; besides, the feature field can effectively reflect spatial feature information of each pixel, describing the original image data, denoted as  $X = \{X | x_1, x_2, \dots, x_n\}$ . According to the Bayesian posterior probability formula [25], the image segmentation can be expressed as

$$P(Y = y | X = x) = \frac{P(X = x | Y = y)P(Y = y)}{P(X = x)}, \quad (6)$$

where  $P(Y = y | X = x)$  is the posterior probability of the marker field  $Y = y$  under the condition of the given image characteristic  $X = x$ ,  $P(X = x | Y = y)$  is the probability distribution of characteristic field  $X = x$  under the condition of given observation data  $Y = y$ ,  $P(Y = y)$  represents the prior probability, and  $P(X = x)$  is a known constant so it can be ignored. The maximum posterior probability can be approximated:

$$\hat{y} = \arg \max P(Y = y | X = x). \quad (7)$$

During parameter solving process, the image probability calculation actually can be converted into a problem of energy solution, i.e., to minimize the sum of the current mark field energy and the characteristic field energy, namely,

$$E(y, x) = \min (E_Y(y) + E_X(y, x)). \quad (8)$$

In the formula,  $E_Y(y)$  is the marker field energy, and  $E_X(y, x)$  is the characteristic field energy.

Then, the energy potential functions of the marker field can be defined as

$$E_Y(y_I) = \sum_{k=1}^s v(I, k), \quad (9)$$

$$v(I, k) = \begin{cases} f, & y_I \neq y_k, \\ 0, & y_I = y_k, \end{cases} \quad (10)$$

where  $I$  is a centre pixel of the neighborhood and  $K$  is the pixel contained in the neighborhood.

Assuming that the probability distribution of the image is a normal distribution, the mean and variance of  $k$  in each category are  $\mu_k$  and  $\sigma_k$ , respectively;  $E_X(y, x)$  can be expressed as

$$E_X(y, x) = \sum \left( \ln \left( \sqrt{2\pi} \times \sigma_k \right) + \frac{(x - \mu_k)^2}{2\sigma_k^2} \right). \quad (11)$$

Iterative condition modes (ICM) are used to iterate, the

program is finalized, and the final segmentation result is recorded once the termination condition is met.

*3.1.4. Single-Wood Separation, Extraction, and Postprocessing.* In order to obtain the tree information from segmentation results, it is postprocessed by mathematical morphological operators and other methods. The major processing steps [26] are described as follows:

- (1) The segmentation result area is separated: separate the areas containing tree information from the segmentation result and save in TIFF format
- (2) Isolated nodes removal: the primary purpose is to remove the isolated noise information in the segmentation result
- (3) Patchwork elimination: mainly remove the sporadic small fragmented noise in the non-tree areas
- (4) Hole filling: fill the holes in the tree pattern to form a complete closed polygon for further extraction of the tree crown
- (5) Morphological operator processing: basic morphological operators in this study include expansion operator ( $\oplus$ ), corrosion operator ( $\ominus$ ), open operator ( $\circ$ ), and closure operators ( $\bullet$ ) [27]. The principles expressed by assuming binary image  $A$  and structure element  $B$  are shown in equations (12)–(15). The expansion operator can increase the number of pixels on the border; the erosion operator can reduce the number of pixels on the border. The open operator is first corroded and then expanded; the closed operator is first expanded and then corroded

$$A \oplus B = \cup \{A + b | b \in B\}, \quad (12)$$

$$A \ominus B = \cap \{A - b | b \in B\}, \quad (13)$$

$$A \circ B = (A \ominus B) \oplus B, \quad (14)$$

$$A \bullet B = (A \oplus B) \ominus B \quad (15)$$

This work adopts the processing strategy of first corrosion and consequently expansion, i.e., open operators for morphological processing, which could smooth the boundary contour of the pattern, break the narrow connections, and remove the small protrusion on edges.

- (6) Treatment of abnormal segmentation patches: due to the fact that two or more tree spots might be connected together in the segmentation results, which reduces the extraction accuracy. Thus, it is separated based on the rules of patch area and patch roundness

According to the patch area statistics from the processing results, the patches whose area is more than 90% of the total patch area are screened out and recorded as  $A_u$ , and the remaining patches are recorded as  $A_d$ ; then, the patch roundness is calculated, and spots with roundness less than

TABLE 1: The statistical results of DBI.

Number of categories	DBI	
	<i>Pinus sylvestris</i> forest area	<i>Pinus tabulaeformis</i> forest area
2	1.43537	1.44351
3	1.15484	1.18965
4	1.13441	1.02295
5	1.24226	1.08863
6	1.07897	1.18318
7	1.12276	1.14299
8	1.10650	1.23357
9	1.20727	1.18306
10	1.13171	1.14993
11	1.17526	1.11194
12	1.18779	1.10884
13	1.14118	1.09068
14	1.12798	1.07564
15	1.11344	1.07782

0.5 is marked as  $C_d$ , and the remaining patches are marked as  $C_u$ . The  $C_d$  patches were corroded and then combined with  $C_u$  and  $A_d$  to produce final segmentation result.

**3.1.5. Extraction of Individual Standing Trees.** For individual standing tree identification, based on the principle of geometric image moments, the morphological postprocessing results are calculated to obtain the centroid position of the segmentation pattern, which is used as a marker [28] for the crown position, and the number of markers is the exact number of trees extracted. The calculation formulae are

$$x = \frac{C_{10}}{C_{00}} = \frac{\sum_{i=1}^M \sum_{j=1}^N i \times f(i, j)}{\sum_{i=1}^M \sum_{j=1}^N f(i, j)}, \quad (16)$$

$$y = \frac{C_{01}}{C_{00}} = \frac{\sum_{i=1}^M \sum_{j=1}^N j \times f(i, j)}{\sum_{i=1}^M \sum_{j=1}^N f(i, j)}. \quad (17)$$

$f(i, j)$  is the grey value of the image pixel;  $C_{10}$  and  $C_{01}$  are the first-order moments of the image, representing the cumulative sum of the product of  $x$ -coordinate value and pixel grey value of the patch area and the cumulative sum of the product of the  $y$ -coordinate value of the patch area and the pixel grey value;  $C_{00}$  is the zero-order moment of the image, which represents the cumulative sum of pixel grey values in the pattern area. Additionally, by vectorizing the final morphological postprocessing results, the corresponding single-tree canopy range is obtained. Note that now, moment technology has been widely used in image retrieval and recognition, image matching, image reconstruction, digital compression, digital watermarking, and moving image sequence analysis. Common moment descriptors can be divided into the following categories: geometric moment, orthogonal moment, complex moment, and rotational moment.

**3.2. Verification and Evaluation of Individual Standing Tree Information Extraction.** When evaluating the methods for extracting information from individual trees, relevant reference data, including the position of individual trees and the range of individual tree crowns, should be considered for accuracy evaluation. Due to the lack of measured reference data, this study employs visual interpretation instead of reference data for accuracy verification.

**3.2.1. Evaluation of Recognition Accuracy of Individual Standing Trees.** The evaluation accuracy indicators of individual standing tree identification are basically overall accuracy, commission error, and omission error [29]. The number of trees obtained by artificial recognition is used as the reference tree  $N_r$ . The number of trees extracted is recorded as  $N_a$ , and the number of trees extracted correctly is  $N_c$ . The overall accuracy refers to the percentage of the number of correctly extracted trees to the number of reference trees, see equation (18). Commission error refers to the percentage of the difference between the number of reference trees and correctly extracted trees in the number of reference trees, as shown in equation (19). Omission error refers to the percentage of the difference between the number of extracted trees and correctly extracted trees in the number of reference trees, see formula (20). In this paper, all the experiments have gone through a lot of experiments, and we have carried out three experiments, respectively, under each working condition, and then obtained the final results by using the average value technology, so the results are very effective.

$$\text{Overall accuracy} = \text{OA} = \frac{N_c}{N_r} \times 100\%, \quad (18)$$

$$\text{Commission error} = \text{CE} = \frac{N_r - N_c}{N_r} \times 100\%, \quad (19)$$

$$\text{Omission error} = \text{OE} = \frac{N_a - N_c}{N_a} \times 100\%. \quad (20)$$

**3.2.2. Evaluation of Accuracy of Single-Tree Crown Extraction.** The range of the canopy was extracted manually by professional interpreters as a reference for verifying the results of single-tree crown extraction. This article quantitatively evaluates the accuracy of single-tree crown extraction based on the following criteria:

- (1) Matching: the extracted single-tree canopy overlaps with the reference canopy by more than 50%, and the centre point of the reference canopy is located in the extracted canopy
- (2) Merge: the extracted single tree canopy contains two or more reference canopies, and the centre of the reference crown is also included in the extracted crown
- (3) Separation: more than 50% of the reference tree canopy is occupied by one or more single-tree canopies
- (4) Loss: the overlap between the reference canopy and the extracted single tree canopy is less than 50%



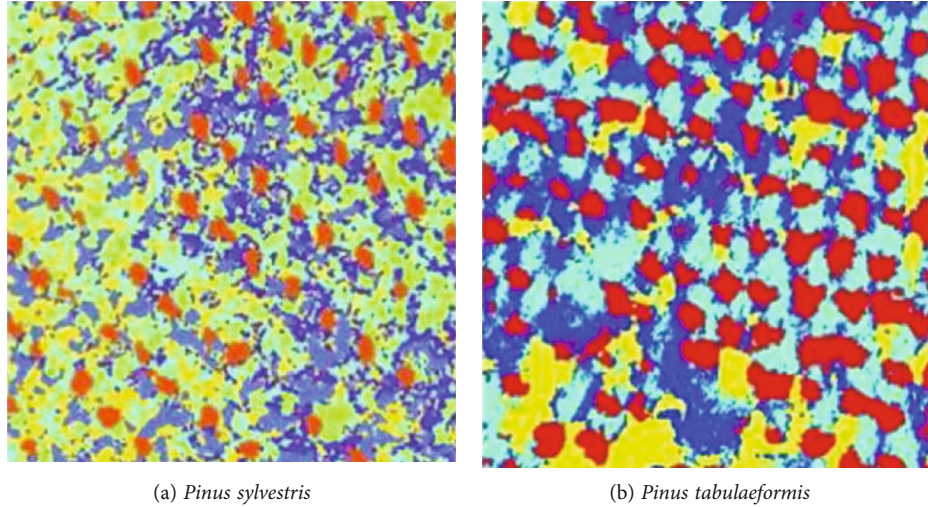


FIGURE 4: The segmented areas of the algorithm of the study.

TABLE 2: Segmentation parameters.

Experimental area	Band weight	Segmentation scale	Shape factor	Compactness factor
<i>Pinus sylvestris</i> area	1 1 1	20	0.8	0.2
<i>Pinus tabulaeformis</i> area	1 1 1	60	0.65	0.5

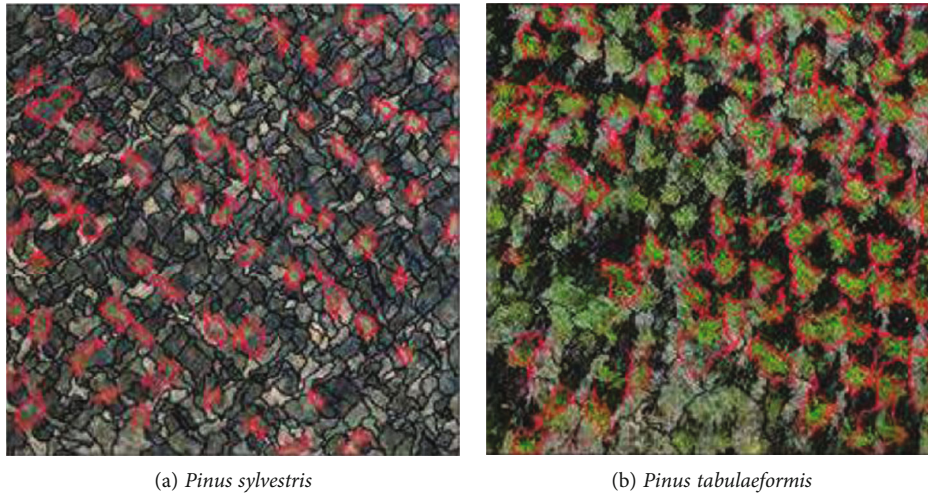


FIGURE 5: Segmentation results of object-oriented method.

[30], and the centre of the reference canopy does not belong to any extracted single-tree crown

## 4. Results and Discussion

**4.1. DBI Optimization Results.** Based on the UAV remote sensing image data of the *Pinus sylvestris* forest and the *Pinus tabulaeformis* forest, the corresponding DBI index is calculated through the program for different preset classification numbers. The statistical results are shown in Table 1.

It can be seen that the minimum DBI of the *Pinus sylvestris* area is 1.07897, and the number of classifications is 6, and it is taken as the optimal number of classifications;

The minimum value of DBI is 1.02295 of the *Pinus tabulaeformis* forest, and the number of classifications is 4, which is regarded as the optimal number of classifications.

**4.2. Segmentation Result Comparison and Analysis.** The segmentation results of UAV remote sensing image of *Pinus sylvestris* forest and *Pinus tabulaeformis* forest are greyscaled image. The UAV remote sensing images of *Pinus sylvestris* forest are divided into 6 and 4 categories, respectively. It is difficult to distinguish different divided regions in the grey-scale image visually; therefore, in order to facilitate observation, the colours of the different segmented areas are, respectively, matched, see Figure 4.

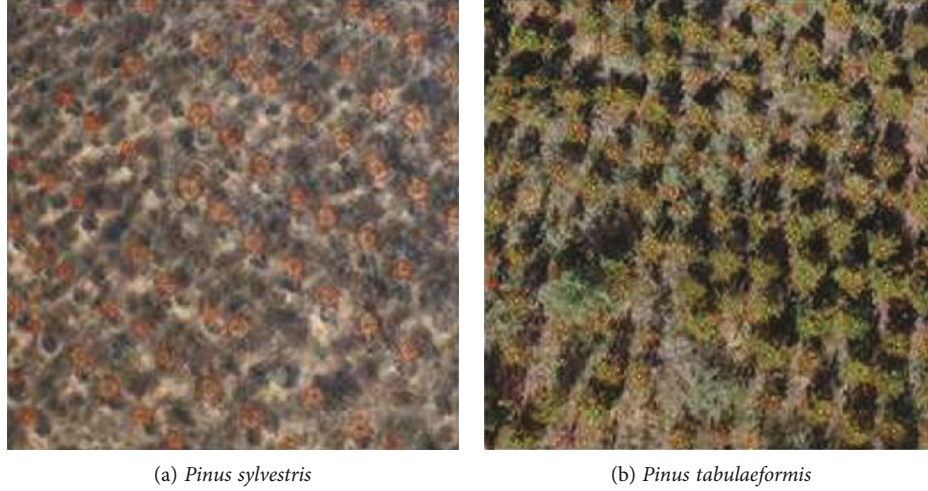


FIGURE 6: The results of individual standing tree identification of *Pinus sylvestris* and *Pinus tabulaeformis*.

TABLE 3: The accuracy evaluation of the individual standing tree recognition.

Experimental area	$N_r$	$N_a$	$N_c$	OA (%)	CE (%)	OE (%)
<i>Pinus sylvestris</i> forest area	69	67	66	95.66	4.37	2.89
<i>Pinus tabulaeformis</i> forest area	105	103	94	89.55	10.39	8.81

Since the UAV remote sensing image in this paper has a high spatial resolution, a transparent shadow area can be seen from the original image. It mainly consists of two parts. One part is self-shadow, and the shadow area comes from the branches or leaves in the crown; the other part is cast shadow, and the shadow area caused by the sun's rays on the ground or other trees. From the segmentation results, it can be seen that the segmentation algorithm in this paper can suppress or eliminate the shadows of trees to a certain extent. Moreover, casting shadows can also be better distinguished and separated. It shows that this algorithm has better robustness to the shadow part.

In order to further verify the segmentation effect of this article, the object-oriented segmentation method is applied to segment the experimental images. Based on the software platform eCognition Developer-10.2, a multiscale segmentation, the segmentation parameters include image layer weights, segmentation scale, shape factor, and compactness factor [31]. The weight of the band can be determined according to the importance of the band; the segmentation scale determines the size of the segmented object and the quality of the segmentation; the shape factor determines the proportion of the spectral value relative to the shape during the segmentation; the compactness factor determines the shape characteristics of generated object.

After several trials of the segmentation process, the optimal segmentation parameters of different UAV remote sensing images are determined, see Table 2. The final segmentation results are shown in Figure 5. In order to compare the segmen-

tation effects, the tree objects in the segmentation area are highlighted.

By comparing the segmentation results of the algorithm in this paper with that of the object-oriented method, the algorithm of this paper has a better outcome of the restoration for original tree contours. Due to the complexity of the features in the original image, when multiscale segmentation [32] is performed with different segmentation parameters, the segmentation effect is usually better for some particularly segmented areas, and it is always hard to achieve a balanced segmentation effect for the whole.

**4.3. Individual Standing Tree Identification and Accuracy Evaluation.** The results of individual standing tree identification by UAV remote sensing images of *Pinus sylvestris* forest and *Pinus tabulaeformis* forest are shown in Figure 6. The yellow dots “•” are the reference tree position for artificial visual interpretation, and the red signs “+” are the automatic extraction of the tree position. To improve visual observation, a red circle equivalent to the area of the extracted tree crown is used as marking for the fitting tree range. If the position of the reference tree for manual visual interpretation is within the range of the fitted range, it is regarded as a correct extraction; otherwise, it is regarded as an incorrect one.

The accuracy evaluation of the individual standing tree recognition results is shown in Table 3. For relatively sparse *Pinus sylvestris* forest, the overall accuracy of individual standing tree identification reached 95.66%, the commission error was 4.37%, and the omission error was 2.89%; for the densely growing *Pinus sylvestris* forest area, the overall accuracy of individual standing wood identification reached 89.55%, commission error was 10.39%, and omission error was 8.81%.

**4.4. Single-Tree Crown Extraction and Accuracy Evaluation.** The single-tree crown extraction results of *Pinus sylvestris* and *Pinus tabulaeformis* forest areas are shown in Figure 7. The yellow polygons are the reference crown range obtained

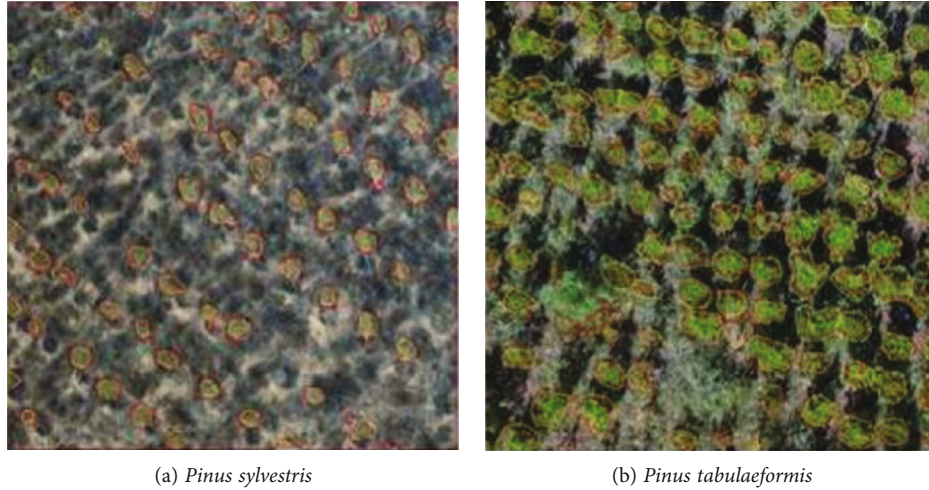
FIGURE 7: The single-tree crown extraction results of *Pinus sylvestris* and *Pinus tabulaeformis*.

TABLE 4: Evaluation of the accuracy of single tree crown extraction.

Experimental area	Reference canopy	Match	Merge	Separation	Loss	Extraction accuracy (%)
<i>Pinus sylvestris</i> forest	69	66	0	1	3	95.66
<i>Pinus tabulaeformis</i> forest	105	86	8	0	9	81.92

by manual visual interpretation, and the red polygons are the automatic extraction range of single-tree crown.

The accuracy evaluation for individual tree crown extraction is shown in Table 4. The number of reference tree crowns for visual interpretation of *Pinus sylvestris* forest is 69, the matching number is 66, the number of separation is 1, the missing number is 3, and the corresponding extraction accuracy is 95.66%; the number of reference tree crowns for visual interpretation of *Pinus tabulaeformis* forest is 105, the matching number is 82, the combined number is 10, the missing number is 8, and the corresponding extraction accuracy is 81.92%.

The method in this paper has higher accuracy for single-tree crown extraction in sparse forest areas, but relatively low extraction accuracy for denser forest areas. The main reason is that the vegetation types of the selected *Pinus tabulaeformis* forest experimental area are complex and diverse, and a large number of shrubs and trees grow together, which are difficult to differentiate. The method in this paper is mainly based on the spectral feature information of ground and mathematical geometry principles to extract single-tree information; the overlapping growth of branches between trees or between trees and shrubs cannot be eliminated [33], which limits the extraction accuracy to a certain extent.

## 5. Conclusions

Aiming for the high spatial resolution UAV remote sensing image data application, a new method for extracting the standing tree information of forest land is proposed in this paper. By comparing with the object-oriented method, it is found that the extraction effect of this method is better.

Through the accuracy verification and evaluation of the extraction results, it is found that the test images for the *Pinus sylvestris* forest area and the *Pinus tabulaeformis* forest area have reached a higher level of accuracy, which shows the method of this paper is both feasible and effective and renders specific application values in remote sensing fields.

The advantages of the method mainly lie in the fact that it does not require a large amount of manual intervention and input of prior knowledge and also avoids the inefficiency of traditional manual extraction. It dramatically improves the degree of automation and universal utilization for the method. Furthermore, it makes full use of the spatial correlation between the grey value of the image pixel and the pixel. It combines the filtering rules based on area and roundness in the postprocessing process, which dramatically improves the extraction effect, accuracy, and shadow inhibition. Discussions were carried out from two aspects (research image data and traditional theoretical method application), focusing on the analysis of the difficulties and limitations of traditional high spatial resolution remote sensing image data in practical application. However, this article only uses high spatial resolution UAV remote sensing image data to extract the information of a single tree. Therefore, how to combine multisource remote sensing data to achieve complementary advantages between data to compensate for the defects of UAV image data and to improve further extraction accuracy of forest information needs further study. In addition, because the algorithm proposed in this paper is more complex, although it can run on the computer with graphics card, it is difficult to guarantee the real-time performance. In this case, this is the place where we need to break through in our next work.

## Data Availability

The data underlying the results presented in the study are available within the manuscript.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This study was financially supported by the Science and Technology Research Project of Chongqing Education Commission (Grant No. KJQN201803402).

## References

- [1] M. Bosse, R. Zlot, and P. Flick, "Zebedee: design of a spring-mounted 3-D range sensor with application to mobile mapping," *IEEE Transactions on Robotics*, vol. 28, no. 5, pp. 1104–1119, 2012.
- [2] M. J. Cracknell and A. M. Reading, "Geological mapping using remote sensing data: a comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information," *Computational Geosciences*, vol. 63, pp. 22–33, 2014.
- [3] C. Cabo, S. del Pozo, P. Rodríguez-Gonzálvez, C. Ordóñez, and D. González-Aguilera, "Comparing terrestrial laser scanning (TLS) and wearable laser scanning (WLS) for individual tree modeling at plot level," *Remote Sensing*, vol. 10, no. 4, p. 540, 2018.
- [4] S. Durrieu, C. Vega, M. Bouvier, F. Gosselin, J. P. Renaud, and L. Saint-André, "Optical remote sensing of tree and stand heights," in *Land Resources Monitoring, Modeling, and Mapping with Remote Sensing*, pp. 449–485, CRC Press, 2015.
- [5] Z. Jin, G. Azzari, C. You et al., "Smallholder maize area and yield mapping at national scales with Google earth engine," *Remote Sensing of Environment*, vol. 228, pp. 115–128, 2019.
- [6] J. R. Lehmann, F. Nieberding, T. Prinz, and C. Knoth, "Analysis of unmanned aerial system-based CIR images in forestry—a new perspective to monitor pest infestation levels," *Forests*, vol. 6, no. 12, pp. 594–612, 2015.
- [7] D. Panagiotidis, A. Abdollahnejad, P. Surový, and V. Chiteculo, "Determining tree height and crown diameter from high-resolution UAV imagery," *International Journal of Remote Sensing*, vol. 38, no. 8–10, pp. 2392–2410, 2017.
- [8] S. Bonnet, J. Lisein, and P. Lejeune, "Comparison of UAS photogrammetric products for tree detection and characterization of coniferous stands," *International Journal of Remote Sensing*, vol. 38, no. 19, pp. 5310–5337, 2017.
- [9] J. Estornell, B. Velázquez-Martí, I. López-Cortés, D. Salazar, and A. Fernández-Sarría, "Estimation of wood volume and height of olive tree plantations using airborne discrete-return LiDAR data," *GIScience & Remote Sensing*, vol. 51, no. 1, pp. 17–29, 2014.
- [10] S. Ganz, Y. Käber, and P. Adler, "Measuring tree height with remote sensing—a comparison of photogrammetric and LiDAR data with different field measurements," *Forests*, vol. 10, no. 8, p. 694, 2019.
- [11] P. Cao, D. Zhao, and O. Zaiane, "An optimized cost-sensitive SVM for imbalanced data learning," in *Advances in Knowledge Discovery and Data Mining*, vol. 7819, pp. 280–292, Springer, Berlin, Heidelberg, 2013.
- [12] S. Krause, T. G. M. Sanders, J. P. Mund, and K. Greve, "UAV-based photogrammetric tree height measurement for intensive forest monitoring," *Remote Sensing*, vol. 11, no. 7, p. 758, 2019.
- [13] Z. Jin, G. Azzari, M. Burke, S. Aston, and D. Lobell, "Mapping smallholder yield heterogeneity at multiple scales in eastern Africa," *Remote Sensing*, vol. 9, no. 9, p. 931, 2017.
- [14] S. Madonsela, M. A. Cho, R. Mathieu et al., "Multi-phenology WorldView-2 imagery improves remote sensing of savannah tree species," *International Journal of Applied Earth Observation and Geoinformation*, vol. 58, pp. 65–73, 2017.
- [15] A. Lausch, S. Erasmi, D. J. King, P. Magdon, and M. Heurich, "Understanding forest health with remote sensing—part II—a review of approaches and data models," *Remote Sensing*, vol. 9, no. 2, p. 129, 2017.
- [16] C. S. Daughtry, C. Walthall, M. Kim, E. B. de Colstoun, and J. McMurtrey, "Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance," *Remote Sensing of Environment*, vol. 74, no. 2, pp. 229–239, 2000.
- [17] J. A. Long and R. L. Lawrence, "Mapping percent tree mortality due to mountain pine beetle damage," *Forest Science*, vol. 62, no. 4, pp. 392–402, 2016.
- [18] J. Lisein, M. Pierrot-Deseilligny, S. Bonnet, and P. Lejeune, "A photogrammetric workflow for the creation of a forest canopy height model from small unmanned aerial system imagery," *Forests*, vol. 4, no. 4, pp. 922–944, 2013.
- [19] J. C. White, N. C. Coops, M. A. Wulder, M. Vastaranta, T. Hilker, and P. Tompalski, "Remote sensing technologies for enhancing forest inventories: a review," *Canadian Journal of Remote Sensing*, vol. 42, no. 5, pp. 619–641, 2016.
- [20] H. Huang, S. He, and C. Chen, "Leaf abundance affects tree height estimation derived from UAV images," *Forests*, vol. 10, no. 10, p. 931, 2019.
- [21] J. Dempewolf, J. Nagol, S. Hein, C. Thiel, and R. Zimmermann, "Measurement of Within-Season Tree Height Growth in a Mixed Forest Stand Using UAV Imagery," *Forests*, vol. 8, no. 7, p. 231, 2017.
- [22] K. V. Tubby and J. F. Webber, "Pests and diseases threatening urban trees under a changing climate," *Forestry: An International Journal of Forest Research*, vol. 83, no. 4, pp. 451–459, 2010.
- [23] J. Lordan, M. Pascual, F. Fonseca et al., "An image-based method to study the fruit tree canopy and the pruning biomass production in a peach orchard," *HortScience*, vol. 50, no. 12, pp. 1809–1817, 2015.
- [24] P. Tittonell, A. Muriuki, C. J. Klapwijk, K. D. Shepherd, R. Coe, and B. Vanlauwe, "Soil heterogeneity and soil fertility gradients in smallholder farms of the east African highlands," *Soil Science Society of America Journal*, vol. 77, no. 2, pp. 525–538, 2013.
- [25] C. Z. Espinoza, L. R. Khot, S. Sankaran, and P. W. Jacoby, "High resolution multispectral and thermal remote sensing-based water stress assessment in subsurface irrigated grapevines," *Remote Sensing*, vol. 9, no. 9, p. 961, 2017.
- [26] F. Giannetti, N. Puletti, V. Quatrini et al., "Integrating Terrestrial and Airborne Laser Scanning for the Assessment of Single-Tree Attributes in Mediterranean Forest Stands," *European Journal of Remote Sensing*, vol. 51, no. 1, pp. 795–807, 2018.

- [27] J. Li, B. Yang, Y. Cong, L. Cao, X. Fu, and Z. Dong, "3D forest mapping using a low-cost UAV laser scanning system: investigation and comparison," *Remote Sensing*, vol. 11, no. 6, p. 717, 2019.
- [28] F. Kitahara, N. Mizoue, and S. Yoshida, "Effects of training for inexperienced surveyors on data quality of tree diameter and height measurements," *Silva Fennica*, vol. 44, pp. 657–667, 2010.
- [29] M. Despotovic, V. Nedic, D. Despotovic, and S. Cvetanovic, "Evaluation of empirical models for predicting monthly mean horizontal diffuse solar radiation," *Renewable and Sustainable Energy Reviews*, vol. 56, pp. 246–260, 2016.
- [30] C. Qian, H. Liu, J. Tang et al., "Assessment of forest structure using two UAV techniques: a comparison of airborne laser scanning and structure from motion (SfM) point clouds," *Forests*, vol. 7, no. 12, p. 62, 2016.
- [31] L. Wallace, A. Lucieer, C. Watson, and D. Turner, "Development of a UAV-LiDAR system with application to forest inventory," *Remote sensing*, vol. 4, no. 6, pp. 1519–1543, 2012.
- [32] R. A. Díaz-Varela, R. de la Rosa, L. León, and P. J. Zarco-Tejada, "High-resolution airborne UAV imagery to assess olive tree crown parameters using 3D Photo reconstruction: application in breeding trials," *Remote Sensing*, vol. 7, no. 4, pp. 4213–4232, 2015.
- [33] R. Perko, H. Raggam, K. Gutjahr, and M. Schardt, "Advanced DTM generation from very high-resolution satellite stereo images," in *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, vol. 2, Munich, Germany, 2015.

## Research Article

# Multiscale Bidirectional Input Convolutional and Deep Neural Network for Human Activity Recognition

Yishu Qiu, Lanliang Lin , Lvqing Yang , Dingzhao Li, Runhan Song, Gengchen Xu, and Shaoqin Shen

*School of Informatics, Xiamen University, Xiamen, Fujian Province 361005, China*

Correspondence should be addressed to Lvqing Yang; [lqyang@xmu.edu.cn](mailto:lqyang@xmu.edu.cn)

Received 15 July 2021; Revised 9 August 2021; Accepted 23 August 2021; Published 18 October 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Yishu Qiu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this paper, we proposed a multiscale and bidirectional input model based on convolutional neural network and deep neural network, named MBCDNN. In order to solve the problem of inconsistent activity segments, a multiscale input module is constructed to make up for the noise caused by filling. In order to solve the problem that single input is not enough to extract features from original data, we propose to manually design aggregation features combined with forward sequence and reverse sequence and use five cross-validation and stratified sampling to enhance the generalization ability of the model. According to the particularity of the task, we design an evaluation index combined with scene and action weight, which enriches the learning ability of the model to a great extent. In the 19 kinds of activity data based on scene+action, the accuracy and robustness are significantly improved, which is better than other mainstream traditional methods.

## 1. Introduction

The research of using sensors for human activity recognition existed as early as 30 years ago [1, 2]. With the advent of industry 4.0 era, new opportunities and challenges are filled. Human activity recognition research has attracted much attention due to its advantages in intelligent monitoring system, medical care system [3], virtual reality exchange, smart homes [4], anomaly detection [5], and other fields, as well as the ability to provide personalized support and interconnection for different fields. At present, human activity recognition is mainly realized by two ways: one is through indoor and outdoor sensors, and the other is some wearable devices [6–8]. The former is limited by the need to be placed in a fixed location, and the inference of activity completely depends on the user's interaction with these devices. For example, if the user is not within the sensor range or the object moves freely in the scene to introduce varying degrees of occlusion, the activity cannot be recognized. Secondly, the environment is dynamic and complex, such as the weather and sunlight in the background, which also increases the difficulty of recognition. The latter also has many defects, such as high cost and inconvenient carrying.

Smart phones have many advantages in the field of human activity recognition [9]. Due to its small size and convenient portability, the built-in sensors are becoming more and more diverse, and specific types of activities can be effectively classified through the information of multiple sensors. For example, the built-in accelerometers of smart phones [10, 11] can describe human actions, such as standing, walking, and running [12, 13]. Similarly, by collecting audio information from the phone microphone [14], the user's activities can be identified, such as listening to music, speaking, and sleeping [15, 16], running rhythm can be monitored [17], and user respiratory symptoms are related to sound, such as sneezing or coughing [15]. Users' emotions can also be inferred from various sensor data, including Wi-Fi, accelerometer, compass, and GPS [15, 18–20]. The activity identification system based on mobile devices can be perceived from personal perception to group in multiscale [21]. Generally speaking, the quality of data collected by each built-in sensor of a smart phone is also different. For example, the gyroscope of smart phone can sense the change of the movement direction of the person holding the phone, while the acceleration sensor of smart phones can reflect

the speed change of the person holding the phone. Therefore, by fusing the real-time data obtained by the various sensors of the smart phone, the final data obtained has a certain heterogeneity. These data can be widely used in human activity recognition and have broad market and social value in health care, smart home, financial fraud, and other scenarios. At the same time, the user does not need to carry additional equipment with sensors in the process of collecting data. In terms of human activity recognition, using smart phones as research equipment has become the preferred equipment for researchers.

We study the indoor and outdoor human activity recognition, redefine activity as the combination of scene and action, and achieve real-time monitoring of users' indoor and outdoor activity through smart phones.

- (1) From the study of recognition simple activity to recognition complex activity, activity is defined as the combination of scene and action, which has 19 different activities. The aggregation feature is designed manually to help automatic feature extraction and realize more abundant feature information extraction
- (2) A multiscale bidirectional sequence fusion model (MBCDNN) is designed to recognition complex activity. The experiment result shows that the model has higher advantages than the current popular network model

## 2. Related Work

**2.1. Dataset Collected by Smart Phone.** In the research of human activity recognition, the data collected by a smart phone through sensors is particularly important. Some of these benchmark datasets have been released to the public. In studies, such as Roggen et al.'s [22], using 72 environmental and body sensors, a set of daily activities were recorded in the sensor rich environment. Similarly, the other researchers have provided dataset, such as Tapia et al. [23] and Hasegawa[24]. In 2012, WISDM Lab released wisdm dataset [25], the device for collecting data is an Android smartphone, and the number of users participating in data collection is 29. Users participating in data collection are required to put the smartphone in the front trouser leg pocket, which is more in line with the real-life scene. Users need to complete a series of actions, including walking, jogging, going up and down the stairs, standing, and sitting. In 2013, the University of Genoa provided UCI-HAR dataset [26] to record the daily activity of each user, including standing, sitting, walking, going up and down the stairs, and lying down. With the help of the acceleration sensor and gyroscope of the smart phone, the data collection device collects 30 users, aged from 19 to 48, and each user is instructed to wear a smart phone at the waist. Each user needs to carry out two experiments. In the first experiment, the smart phone is worn on the left side of the waist, and in the second experiment, the user can freely choose the position. In 2016, Vavoulas and others offered MobiAct dataset [27], Mobiact data is an extension of the MobiFall data set published in

2014 and was initially created with fall detection in mind. This dataset contains four different types of falls and nine daily activities, such as walking, standing, and going up and down the stairs. The number of users collected is 57, including 42 men and 15 women. The users are between 20 and 47 years old, including more than 2500 experiments, all from the collection of smart phone sensors. In the same year, Hnoohom and others provided the UniMiBSHAR dataset [28]. In the authors' opinion, the data collected by smart phone sensors are rarely public, and public data often contain samples of users with too similar characteristics and lack of specific information. They proposed a new dataset, which collected daily activities of users mainly including simple activity such as walking, standing, running, and sitting down and complex activity such as washing dishes, combing hair, and preparing sandwiches. The dataset contains 11771 human activities and 30 users aged between 18 and 60, of which 24 are women and 6 are men. It is worth noting that it includes more elderly people. In 2019, Beijing University of Posts and Telecommunications provided the Sanitation dataset [29]. The user receiving the acquisition needs to wear a smart watch at the wrist. The user selected according to the demand is the sanitation worker. According to the research requirement, sanitation workers were invited as a user, in which 7 kinds of daily life action data that were collected included walking, running, sweeping with a big broom and sweeping with a small broom, cleaning, and taking out garbage.

**2.2. HAR Based on AI.** Reference [30] collected data from 10 volunteers (4 women and 6 men) through four sensors of smart phones. The volunteers are between 24 and 30 years old. Volunteers are required to carry smart phones to complete the specified actions, standing, walking, walking slowly, go upstairs and downstairs and cycling. In order to eliminate errors, discard the data of the first 2 seconds and the second after the volunteer starts the action, and select 50050 data for each action. They processed data using median filtering, data normalization, sliding window segmentation, feature selection, and optimal feature subset selection are performed on the data. Also, they used the idea of fusion. The traditional random forest (RF) [31], support vector machine (SVM) [32],  $k$ -nearest neighbor (KNN) [33], and naive Bayes classification algorithm (NBC) [34] are fused. The accuracy was 99%, which is 6% higher than that of single model

With the development of artificial intelligence, deep learning has become a favorite choice of researchers. Reference [35] used smart phone sensors to collect 6 different activities from 30 volunteers, including walking, going upstairs and downstairs, standing, lying flat, and sitting. The collected data are preprocessed by noise filter, and 2.56 seconds are used as sliding window. A total of 7352 training data and 2947 test data are divided. Three design optimization schemes were proposed to improve the accuracy of human activity recognition: (1) Scheme 1—combining three sensor data; (2) Scheme 2—based on three-dimensional convolution; and (3) Scheme 3—based on difference optimization convolution kernel. By comparing the results of the experimental scheme, it was found that the accuracy of the first scheme is the highest, with an average accuracy of 97.50%.

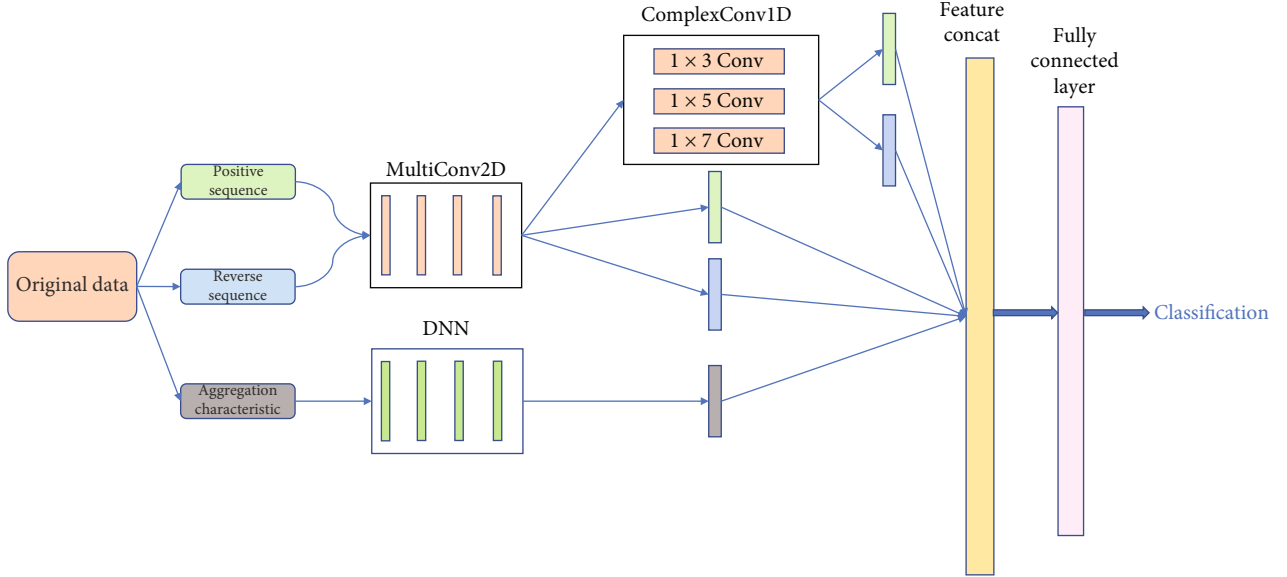


FIGURE 1: MBCDNN model.

Due to the complexity of human activities, there are few researches on excessive activities in the past, such as sitting-standing. Reference [36] proposed a hierarchical recognition model based on support vector machine and random forest, studying 4 kinds of excessive activities and 10 kinds of daily life activities. The data preprocessing adopted filtering denoising and acceleration separation, and the feature extraction adopted the time domain and frequency domain of features, in which the frequency domain uses a total of 118 features, such as Fourier realization, mean, variance, and standard deviation. In the self-collected data, support vector machine and random forest hierarchical recognition model achieved an average accuracy of 98.1%. LSTM is good at processing time series data. The author used bidirectional LSTM network for activity recognition research and achieved 94.1%, 98.5%, and 98.8% accuracy in UCI-HAR, WISDM, and self-collected data, respectively.

### 3. Method

**3.1. MBCDNN Model.** Human activity data collected based on scenes+actions are more complex. Since the collected data is time series data, we send the data of the forward sequence and the reverse sequence to the MultiConv2D module for training. Due to the different activity duration, the collected data length is not consistent. Different experiments in the past have segmented the data, trying to find the best segmentation length [15, 37], but in order to ensure the consistency of the sequence fragments in the division process, the filling strategy is adopted. The obvious problem is that the noise interference is increased. Therefore, we introduce the activity convolution module ComplexConv1D, which uses a series of designed one-dimensional convolution kernels to perform convolution operations on forward and reverse sequences to extract diversified features; the interference caused by noise is largely reduced. In addition to the forward and reverse sequences as input to the model, we also added a series of aggregation

features, including three types of features: time domain, frequency domain, and time-frequency domain, which perform feature extraction on the original data, these extracted aggregation features are input into the deep neural network (DNN) module, and this is verified to be effective. Our MBCDNN model is composed of MultiConv2D and ComplexConv1D and DNN modules. Finally, the output results of all modules are connected, and the classification result is obtained through the fully connected layer. The overall structure of the model is shown in Figure 1.

**3.1.1. MultiConv2D and ComplexConv1D.** The task of the MultiConv2D module is to receive the input of the forward sequence and the reverse sequence, followed by four convolutional layers (Conv2D); the kernel size is  $3 \times 3$ ; the number of convolution channels is 64, 128, 256, and 512, respectively; the activation function is ReLU. Each layer of convolution passes through the Batch Normalization (BN) layer, then passes through the pooling layer. This module extracts the feature of actions and scenes through a series of convolution operations. The formula for two-dimensional convolution is as follows:

$$\text{conv}_{x,y} = f \left( \sum_i^{k \times k} w_i \times v_i \right), \quad (1)$$

where  $x, y$  are the spatial coordinates  $(x, y)$  of the input data sample,  $f$  is the activation function, the weight of the convolution kernel is  $w$ , the size of the convolution kernel is  $k \times k$ , and the sample data value is  $v$ . The convolution process is the sum of the inner product of the weight of the convolution kernel sliding on the value corresponding to the sample data.

The task of the ComplexConv1D module is to extract the features of MultiConv2D from multiple angles when the sequence length is inconsistent, so as to reduce the impact of noise on training. When the input vector dimension is



TABLE 1: MultiConv2D & ComplexConv1D network parameters. Each layer of convolution passes through the BN layer; the selected activation function is ReLU.

Layer name	MultiConv2D Kernel size, filters/stride, pad	Layer name	ComplexConv1D Kernel size, filters/stride, pad
Conv_1 MaxPooling	$3 \times 3, 64/1, \text{same } 2 \times 2/2$	Conv1d_1	$1 \times 7, 128/1, \text{same}$
Conv_2	$3 \times 3, 128/1, \text{same}$	Conv1d_2	$1 \times 5, 256/1, \text{same}$
Conv_3	$3 \times 3, 256/1, \text{same}$	Conv1d_3	$1 \times 3, 128/1, \text{same}$
Conv_4	$3 \times 3, 512/1, \text{same}$		GlobalAveragePooling1D
	GlobalAveragePooling2D		

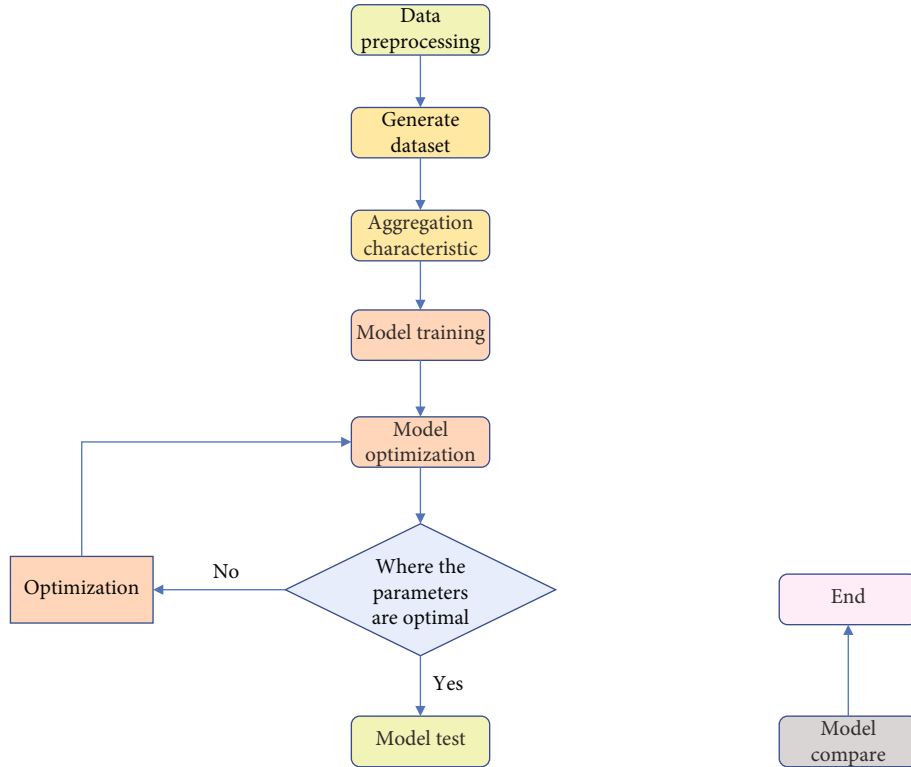


FIGURE 2: Experimental flowchart of MBCDNN.

$(N, C_{in}, L_{in})$ , the output vector dimension is  $(N, C_{out}, L_{out})$ . As shown in

$$L_{out} = \left\lfloor \frac{L_{in} + 2 \times p - d \times (k - 1) - 1}{s} + 1 \right\rfloor, \quad (2)$$

where  $L_{in}, L_{out}$  represent the length of the input and output vectors, respectively,  $p$  represents the padding size,  $d$  represents the distance between the core points, also called à trous algorithm [38],  $k$  represents the size of the convolution kernel, and  $s$  represents the convolution stride. To illustrate the relevant parameters of the model, Table 1 lists the parameters of MultiConv2D and ComplexConv1D in detail. The experimental process of the MBCDNN model is shown in Figure 2. The data processing, generate dataset,

and aggregation characteristic part will be described in this section; the rest will be described in the next section.

**3.2. Data Processing.** The data collected by the smart phone sensor is time series data; for this reason, the research and design collection scenes are divided into 3 categories: walking, standing, sitting, and lying; each scene collects 6 kinds of actions, playing games (mobile game), watching short videos, watching live broadcasts or watching long videos (similar to news broadcasts), browser query or viewing browser content, typing chat or other typing, and other actions (WeChat calls): 6 types of actions. At the same time, an action that is regardless of scene is added, the activity of user A “handing the phone” to user B. We redefine activity as a combination of scene+action, so there are 19 different activities in total. A total of 7292 samples were collected. 80% of the overall data were randomly selected as the training set and 20% as the test set. The sample

TABLE 2: Predefined activity based on smart phone sensors.

Scene	Action	Activity description
Walking/standing/sitting and lying	Playing games	User plays the game in this scene (mobile game)
	Watching short videos	User watches short videos in this scene
	Watching live broadcasts or watching long videos	User watches live broadcast or long videos in this scene (similar to news broadcast)
	Browser query or viewing browser content	User views browser content in this scene
	Typing chat or other typing	User types text in this scene
	Other actions	User makes WeChat calls in this scene, etc.
Regardless of scene	Hand phone	User hands the phone to another user

categories are unbalanced. The proportion of standing playing games is the most, and the proportion of standing brush video is the lowest. Table 2 describes the situation of each activity. Figure 3 shows the proportion of various activities types of collected data.

**3.3. Generated Dataset.** Use the two built-in sensors of the smart phone: gravity sensor and acceleration sensor to collect data. The volunteers who collected the data were composed of 5 men and 5 women, aged between 20 and 55 years. During the collection process, the volunteers can complete 6 types of actions according to each scene (see Table 2 for details). The collected data includes gravitational acceleration, and nongravitational acceleration data and the action segment are divided by the accelerometer count; every 5 seconds is regarded as a segment. The basic attributes of the collected dataset are described in Table 3. A total of 19 actions were collected based on smart phone sensors. A total of 7292 samples were collected. 80% of the overall data were randomly selected as the training set and 20% as the test set.

**3.4. Aggregation Feature.** The data is time series data, which needs to be divided into activity segments. Each activity segment data contains 60 pieces of data, which can be used for feature aggregation. Aggregation time series data features generally select three types of features: time domain, frequency domain, and time-frequency domain; they have been verified to be effective. In time domain, the independent variable is time, and the dependent variable is the change of the signal, which describes the value of the signal at different moments. In frequency domain, the independent variable is the frequency, the dependent variable is the amplitude of the signal, and it describes the spectrogram. Among them, the sample features that can be extracted in the time domain include variance, standard deviation, mean, and skewness. The median frequency, average frequency, and energy spectral density of sample features can be extracted in the frequency domain. The specific calculation formula for sample feature is shown in Table 4.

## 4. Experiment and Analysis

**4.1. Evaluation Index.** Human activity recognition based on scenes and actions cannot simply use the recognition accuracy to evaluate the quality of the model. Since the activity involves actions under the scene, predicting the scene also

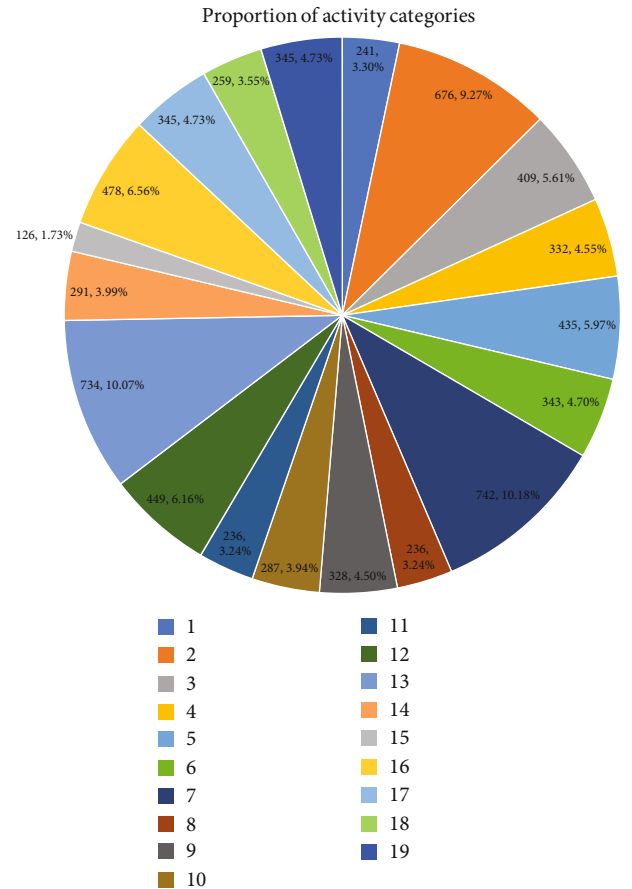


FIGURE 3: Proportion of activity categories. Labels 1 to 6, respectively, represent six types of actions in scene walking: playing games, watching short videos, watching live or long videos, browser query or viewing browser content, typing chat, or other typing and other actions. Labels 7 to 12 represent six types of actions in scene standing. Labels 13 to 18 represent six types of actions in sitting and lying. Label 19 represents the action of delivering the mobile phone alone.

has a certain value. For this reason, we designed an evaluation index that comprehensively considers the weight of the scene and the action: acc\_combo, which is more conducive to the recognition efficiency of the real reaction model. The specific rules of the evaluation index are that if the activity of handing over the mobile phone is predicted to get 1

TABLE 3: Data attribute description table.

Num	Attribute name	Data type	Data description
1	fragment_id	int	Activity fragment ID
2	activity_id	int	Activity ID
3	time_point	float	Acquisition time point (ms)
4	acc <sub>x</sub>	float	X-axis component without gravitational acceleration (m/s <sup>2</sup> )
5	acc <sub>y</sub>	float	Y-axis component without gravitational acceleration (m/s <sup>2</sup> )
6	acc <sub>z</sub>	float	Z-axis component without gravitational acceleration (m/s <sup>2</sup> )
7	acc <sub>xg</sub>	float	X-axis component containing the acceleration of gravity (m/s <sup>2</sup> )
8	acc <sub>yg</sub>	float	Y-axis component containing the acceleration of gravity (m/s <sup>2</sup> )
9	acc <sub>zg</sub>	float	Y-axis component containing the acceleration of gravity (m/s <sup>2</sup> )

TABLE 4: Calculation of time and frequency domain features.

Feature	Calculation formula	Meaning
Mean	$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$	The mean of the activity segment data
Variance	$\sigma = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$	The variance of activity segment data
Standard deviation	$\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$	The standard deviation of the activity segment data
Skewness	$\frac{\sum_{i=1}^n (x_i - \bar{x})^3}{(n-1)\sigma^3}$	The skewness of the activity segment data
Quartile deviation	$Q_3 - Q_1$	The difference between the third quartile and the first quartile
Power spectral density	$P_s(w) = \frac{ F_T(w) ^2}{T}, F_T(w)$ This is the Fourier transform formula	Power in unit frequency band
Median frequency	$\frac{\int_0^{\infty} fP(f)df}{\int_0^{\infty} P(f)df}$	Median of power spectral density
Average frequency	$\int_0^{\infty} P(f)df$	Mean value of power spectral density
Energy spectral density	$G(f) =  F_T(w) ^2$	Signal energy in unit frequency band

point, the prediction error is 0 point. We hope to predict the correct action to get a relatively high positive feedback in the model evaluation. If it is not the activity of handing the phone, the scene+action is completely correct to get 1 point, only the correct scene is predicted to get 1/7 point, and only the correct action is predicted to get 1/3 point; if the scene+action are not correct, the score of acc\_combo is 0 points. The acc\_combo is expressed by

$$\text{acc}_{\text{combo}} = \frac{\sum_{i=1}^N S(i) + A(i)}{N}, \quad (3)$$

where  $N$  represents the total number of prediction samples,  $S$  represents the score of the predicted scene, and  $A$  represents the score of the predicted action.

*4.2. Model Train and Optimization.* Through continuous experimentation to explore, the model is upgraded step by step and finally proposed a Activity Bidirectional input Convolution and Deep Neural Network model: MBDCNN, and the effect is the best among all comparative experiments. The data collected by the smart phone sensor contains components of gravitational acceleration (acc<sub>x</sub>, acc<sub>y</sub>, acc<sub>z</sub>) and components that do not contain gravitational acceleration (acc<sub>xg</sub>, acc<sub>yg</sub>, acc<sub>zg</sub>) (see Table 3). In order to increase the effect of the model, as shown in formulas (4), the components are combined into a vector.

$$\begin{aligned} \text{mod} &= \text{acc}_x^2 + \text{acc}_y^2 + \text{acc}_z^2, \\ \text{mod } g &= \text{acc}_{xg}^2 + \text{acc}_{yg}^2 + \text{acc}_{zg}^2. \end{aligned} \quad (4)$$

TABLE 5: Parameter exploration process.

Version	Model	Input	Modify parameter	acc_combo
1	MultiConv2D	Origin data	Adam \ GlobalMaxPooling2D	0.730
2	MultiConv2D	Origin data	Modify the optimization parameter to rmsprop	0.746
3	MultiConv2D	Origin data	Add the BN layer to the convolution layer	0.786
4	MultiConv2D	Origin data	GlobalAveragePooling2D	0.792
5	MultiConv2D	Forward and reverse sequence	—	0.812

TABLE 6: Model optimization process.

Model	Input	Optimized content	acc_combo
MultiConv2D+DNN	Forward, reverse, and aggregation feature	Added reverse sequence and aggregation features	0.846
MultiConv2D+DNN	Forward, reverse, and aggregation feature	Data enhance	0.862
MBCDNN	Forward, reverse, and aggregation feature	Activity module	<b>0.887</b>

TABLE 7: Comparison results of different model.

Model	Input	acc_combo
Lightgbm	Aggregation feature data	0.774
DNN	Aggregation feature data	0.713
MCNN	Forward and reverse sequence	0.834
CNN	Forward and reverse sequence	0.812
ResNet	Forward sequence	0.828
VGG19	Forward sequence	0.435
Bidirectional LSTM	Forward and reverse sequence	0.708
MBCDNN	Forward, reverse and aggregation feature	0.887

Take the component and mod, mod  $g$  as the input data of the model. To reduce model deviation and overfitting problems, 5-fold cross-validation and stratified sampling are used to ensure that the proportions of different categories in each compromise are equal. This section takes the MultiConv2D module in the MBCDNN model as a baseline. The optimization methods of the model include RMSProp, adding the BN layer after the convolutional layer, GlobalAveragePooling2D, and combining the input of the forward sequence and the reverse sequence. Among them, adding a reverse sequence to the input part has the best effect. The analysis of the reasons shows that the diversified input facilitates the model to more fully extract the feature of activity. The improvement of the specific experimental model is shown in Table 5. In the preliminary work, it is found that the effect of model fusion is better than the effect of single model. So we optimized the model from the following aspects:

- (1) Combine the input of the MultiConv2D module with the input of the DNN module. The input model data is time series data and aggregation feature data. The time series data is input to the MultiConv2D module, and the feature data is input to the DNN module
- (2) For the input part, due to the existence of various scales of activity sequence fragments, a single padding length

cannot be used, and for too long or too short fragments, interception or padding will bring a lot of noise, so we built a activity module: ComplexConv1D, to make up for the impact of noise and enrich the model learning ability

- (3) Data enhancement (noise enhancement, cubic spline interpolation) increases the generalization ability of the model

Table 6 shows the model optimization process.

According to Tables 5 and 6, the effect of multiple models is obviously better than that of single model. Adding reverse sequence and aggregation feature data to the input data has significantly improved the score. Data enhancement is also an important method for score improvement. The addition of activity modules has further improved the score.

*4.3. Model Comparison.* We improve our model based on the idea of convolutional neural network and deep neural network and propose the idea of activity and multi-input; after recognizing 19 kinds of activities, the effect of the algorithm model in this paper is fully proved. Comparative experiments include ensemble learning, deep learning single model, single input model, and multiple input model. Table 7 shows the experimental comparison results of different models.

It can be seen from Table 7 that the scores of the multi-input models are all above 0.8. The MBCDNN model we proposed has a maximum score of 0.887, explaining that input of multiple conversion methods of data, data enhancement, and model fusion can obtain better scores, which further shows that our model has a good activity recognition effect.

## 5. Conclusions

Aiming at the lack of complex action and different scene existing in human activity recognition, we use a smart phone as the carrier equipment and propose complex human activity recognition based on scene+action, by introducing the forward sequence and reverse sequence, as well as aggregation features to help the model with more activity features. However, as the time series data, a truncated or filling strategy will introduce unnecessary noise; for this reason, we propose a ComplexConv1D module to compensate for the impact of unnecessary noise. At the same time, in order to more comprehensively evaluate the performance of the model under a specific activity, we define an evaluation index that combines the weight of the scene and the action. Through experimental comparison and analysis, the performance of our model has indeed been improved, which proves the effectiveness of our method. After all this job we have done, it still needs a lot of things to do on human activity recognition. We believe there will be a more outstanding work in the future.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported by the Research on big data analysis method of beneficiation and metallurgy industry project (project number BGRIMM-KZSKL-2019-03) and the Fujian province industrial field regional development project (project number 2019H4021).

## References

- [1] B. Fu, N. Damer, F. Kirchbuchner, and A. Kuijper, "Sensing technology for human activity recognition: a comprehensive survey," *IEEE Access*, vol. 8, pp. 83791–83820, 2020.
- [2] F. Demrozi, G. Pravadelli, A. Bihorac, and P. Rashidi, "Human activity recognition using inertial, physiological and environmental sensors: a comprehensive survey," *IEEE Access*, vol. 8, pp. 210816–210836, 2020.
- [3] C. Huang, Y. Zong, J. Chen, W. Liu, J. Lloret, and M. Mukherjee, "A deep segmentation network of stent struts based on IoT for interventional cardiovascular diagnosis," *IEEE Wireless Communications*, vol. 28, no. 3, pp. 36–43, 2021.
- [4] J. Rafferty, C. D. Nugent, J. Liu, and L. Chen, "From activity recognition to intention recognition for assisted living within smart homes," *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 3, pp. 368–379, 2017.
- [5] C. Kunhui, *Design and Research of Monitoring System for Elderly Abnormal Activity in Smart House*, Huazhong University of Science and Technology, 2019.
- [6] X. Hu, J. Dai, Y. Huang et al., "A weakly supervised framework for abnormal behavior detection and localization in crowded scenes," *Neurocomputing*, vol. 383, pp. 270–281, 2020.
- [7] F. J. Ordóñez and D. Roggen, "Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition," *Sensors*, vol. 16, no. 1, p. 115, 2016.
- [8] J. P. Wolff, F. Grützmacher, A. Wellnitz, and C. Haubelt, "Activity recognition using head worn inertial sensors," in *Proceedings of the 5th international Workshop on Sensor-based Activity Recognition and Interaction*, pp. 1–7, 2018.
- [9] A. Lentzas and D. Vrakas, "Non-intrusive human activity recognition and abnormal activity detection on elderly people: a review," *Artificial Intelligence Review*, vol. 53, pp. 1–47, 2020.
- [10] N. S. Suriani and F. A. N. Rashid, "Smartphone sensor accelerometer data for human activity recognition using spiking neural network," *International Journal of Machine Learning and Computing*, vol. 11, no. 4, pp. 298–303, 2021.
- [11] A. E. Minarno, W. A. Kusuma, H. Wibowo, D. R. Akbi, and N. Jawas, "Single triaxial accelerometer-gyroscope classification for human activity recognition," in *2020 8th international conference on information and communication technology (ICoICT)*, pp. 1–5, Yogyakarta, Indonesia, 2020.
- [12] D. Garcia-Gonzalez, D. Rivero, E. Fernandez-Blanco, and M. R. Luaces, "A public domain dataset for real-life human activity recognition using Smartphone sensors," *Sensors*, vol. 20, no. 8, p. 2200, 2020.
- [13] C. Xu, *Human Activity Recognition Using Smart Phones*, 2020.
- [14] W. A. Kusuma, A. E. Minarno, and M. S. Wibowo, "Triaxial accelerometer-based human activity recognition using 1D convolution neural network," in *2020 international workshop on big data and information security (IW BIS)*, pp. 53–58, Depok, Indonesia, 2020.
- [15] A. Nandy, J. Saha, and C. Chowdhury, "Novel features for intensive human activity recognition based on wearable and smart phones sensors," *Microsystem Technologies*, vol. 26, pp. 1–15, 2020.
- [16] H. du, Z. Yu, F. Yi, Z. Wang, Q. Han, and B. Guo, "Recognition of group mobility level and group structure with mobile devices," *IEEE Transactions on Mobile Computing*, vol. 17, no. 4, pp. 884–897, 2018.
- [17] P. Garabelli, S. Stavarakis, and S. Po, "Smartphone-based arrhythmia monitoring," *Current Opinion in Cardiology*, vol. 32, no. 1, pp. 53–57, 2017.
- [18] A. Mottelsonand and K. Hornbæk, "An affect detection technique using mobile commodity sensors in the wild," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 781–792, 2016.
- [19] B. Cao, L. Zheng, C. Zhang et al., "Deepmood: modeling mobile phone typing dynamics for mood detection," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 747–755, 2017.
- [20] X. Zhang, W. Li, X. Chen, and S. Lu, "MoodExplorer," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 4, pp. 1–30, 2018.

- [21] N. D. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, and A. Campbell, "A survey of mobile phone sensing," *IEEE Communications Magazine*, vol. 48, no. 9, pp. 140–150, 2010.
- [22] D. Roggen, A. Calatroni, M. Rossi et al., "Collecting complex activity datasets in highly rich networked sensor environments," in *2010 seventh international conference on networked sensing systems (INSS)*, pp. 233–240, Kassel, Germany, 2010.
- [23] E. M. Tapia, S. S. Intille, L. Lopez, and K. Larson, "The design of a portable kit of wireless sensors for naturalistic data collection," in *International Conference on Pervasive Computing*, pp. 117–134, Berlin, Heidelberg, 2006.
- [24] T. Hasegawa, "Smart phones sensor-based human activity recognition robust to different sampling rates," *IEEE Sensors Journal*, vol. 21, no. 5, pp. 6930–6941, 2020.
- [25] J. Shi, D. Zuo, and Z. Zhang, "An energy-efficient human activity recognition system based on smart phones," in *2020 7th International Conference on Soft Computing & Machine Intelligence (ISCMI)*, pp. 177–181, Stockholm, Sweden, 2020.
- [26] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, "A public domain dataset for human activity recognition using smart phones," *Esann*, vol. 3, p. 3, 2013.
- [27] G. Vavoulas, C. Chatzaki, T. Malliotakis, M. Padiaditis, and M. Tsiknakis, "The MobiAct dataset: recognition of activities of daily living using smart phones," in *International Conference on Information and Communication Technologies for Ageing Well and e-Health*, pp. 143–151, 2016.
- [28] N. Hnoohom, S. Mekruksavanich, and A. Jitpattanakul, "Human activity recognition using triaxial acceleration data from smart phones and ensemble learning," in *2017 13th international conference on signal-image Technology & Internet-Based Systems (SITIS)*, pp. 408–412, Jaipur, India, 2017.
- [29] Z. Yu, *Research on Human Activity Recognition Algorithm Based on Time Series*, Beijing University of Posts and Telecommunications, 2019.
- [30] H. Jian and Y. Jiaxian, "Human activity recognition technology based on smart phone multi-sensor fusion," *Journal of Beijing University of Technology*, vol. 46, no. 11, pp. 14–21, 2020.
- [31] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [32] J. Platt, "Sequential minimal optimization: a fast algorithm for training support vector machines," 1998.
- [33] R. D. Short and K. Fukunaga, "The optimal distance measure for nearest neighbor classification," *IEEE Transactions on Information Theory*, vol. 27, pp. 622–627, 1981.
- [34] P. Domingos and M. Pazzani, "On the optimality of the simple Bayesian classifier under zero-one loss," *Machine Learning*, vol. 29, no. 2, pp. 103–130, 1997.
- [35] J. Xiangyu, "Using CNN to realize human activity recognition based on smart phone sensors," Lanzhou University, 2019.
- [36] P. Yun, X. Peiwen, Z. Huaiyu, and L. Junjie, *Human Activity Recognition Method Based on Inertial Sensors*, Zhejiang University, 2020.
- [37] P. K. Shukla, A. Vijayvargiya, and R. Kumar, "Human activity recognition using accelerometer and gyroscope data from smartphones," in *2020 International Conference on Emerging Trends in Communication, Control and Computing (ICONC3)*, pp. 1–6, Lakshmanarh, India, 2020.
- [38] V. Dumoulin and F. Visin, "A guide to convolution arithmetic for deep learning," 2016, <https://arxiv.org/abs/1603.07285>.

## Research Article

# Research on the Influencing Factors of Film Consumption and Box Office Forecast in the Digital Era: Based on the Perspective of Machine Learning and Model Integration

Qi He and Bin Hu 

*School of Management, Shanghai University of Engineering Science, Shanghai 201620, China*

Correspondence should be addressed to Bin Hu; 201771274@yangtzeu.edu.cn

Received 24 August 2021; Revised 12 September 2021; Accepted 15 September 2021; Published 14 October 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Qi He and Bin Hu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The film industry is one of the core industries of the digital creative industry, which has great positive externalities to the digital creative economy. Movie box office revenue is an important indicator to measure the realization of the market value of movie consumption, and it is also the basic guarantee for the sustainable development of the movie industry. This paper relies on the professional database of the Maoyan movie market to use Python software to collect a total of 830 domestic movie-related consumption characteristic data from 2017 to 2019. In this study, the stacking method in the machine learning ensemble algorithm combines the fivefold crossfolding training method based on distributed random forest, extremely randomized trees, and generalized linear models. The model is good at handling different data types. It has higher fitting and model accuracy in feature mining and model construction, so as to effectively grasp the relevant feature factors affecting movie consumption and accurately predict the future movie box office. Based on the innovative design method of model fusion, the extracted feature vector is used to build a more accurate movie box office prediction model through stacking with a fivefold crossfolding training method. It is aimed at opening the black box that affects the realization of the value of the film content consumption market in the digital age and putting forward corresponding countermeasures and suggestions.

## 1. Introduction

With the continuous development of digital technology, the digital transformation characterized by artificial intelligence [1–3] and big data applications has promoted the continuous evolution of the connotations, boundaries, and forms of creative economy and industrial development. The role of enhancing national competitiveness, promoting the development of industrial integration, and inducing new models and new business forms is increasingly deepening, and the impact on social development is becoming more and more profound. Vigorously promoting digital consumption has become an important driving engine for China to build a new development pattern that focuses on the domestic cycle and domestic and international dual cycles.

The film industry fully embodies the integration of humanities and art and technological innovation, the integration of traditional media and digital media, and the inte-

gration of producers and consumers. In the planning and classification of cultural and creative industries and digital content industries in different countries and regions, it has always been in the core category, which has great positive externalities to the digital economy. Movie products are a typical representative of the development of creative cultural products and digital content. Movie box office revenue is an important indicator to measure the realization of the consumer market value of the movie industry. As of 2019, the Chinese film industry has leaped to the second place in the world in terms of market size and has made important contributions to the economic benefits and social impact of the domestic digital content industry, although the development of the new crown epidemic in 2020 has affected the offline film industry to a certain extent. But at the same time, the reshaping of the film industry by digital transformation has penetrated the entire industry chain, profoundly changing the format and ecosystem of the film industry. The

operation logic of deep integration of technology and creativity is deeply rooted in the hearts of the people. With the gradual development of big data and artificial intelligence, digital technology has penetrated into the entire industry chain of production, distribution, and sales of the film industry, including the algorithm strategy to open the technical support for audiovisual streaming media on the online distribution of movies, as well as the opening of the artificial intelligence system's intervention in film production management such as box office prediction and audience positioning [4]. Since 2020, many film and television groups, including Hollywood giant Warner Bros., have built their own artificial intelligence project management systems, trying to gradually use artificial intelligence-related technologies to evaluate the value of content and main creation, so as to assist the decision-making reference of film distribution strategies [5].

However, the consumption of film products in the digital economy era is affected by multiple factors, and its box office forecasts are more challenging. Although previous studies have conducted a series of empirical analyses using statistical analysis methods and related indicators, the simple use of statistical analysis models is not enough to deconstruct the complex characteristics and structural relationships of film consumption under the new pattern. At present, there is still no method that comprehensively considers the comprehensive characteristics of film consumption in the context of digital transformation to conduct in-depth systematic research, which is insufficient for accurately grasping the characteristics of factors affecting digital content creative consumption and interpreting and predicting the future value of the box office. Therefore, based on the original research, this article systematically analyzes the multidimensional factors affecting film consumption in the digital age, relying on the professional database of the Maoyan film market, and comprehensively using the research methods of big data and machine learning to extract and construct the characteristics of relevant consumption influencing factors. Through model fusion training an innovative and enhanced predictive model, it attempts to build a research framework for the factors affecting film consumption in the context of digital transformation and opens the black box that affects movie box office. The main tasks of this study are as follows:

- (1) Data collection and preprocessing. The data sources of this research mainly come from the famous professional movie website Maoyan professional database, Sina Weibo, IMDB professional database, and WeChat official account platform. The data provided by these platforms were manually screened for unit and text errors, as well as data cleaning of error data, redundant data, and missing data during the data transmission process. A total of 830 movies were indexed
- (2) The stacking method in the machine learning ensemble algorithm combines the fivefold crossfolding training method based on distributed random

forest, extremely randomized trees, and generalized linear models. The model is good at handling different data types. This method has higher fitting and model accuracy in feature mining and model construction, so as to effectively grasp the relevant feature factors affecting movie consumption and accurately predict the future movie box office

## 2. Related Research

*2.1. Research on Influencing Factors of Movie Box Office.* The research on the factors affecting movie box office has a long history. It can be traced back to the 1940s. Early research mainly focused on research techniques [6]. For the first time, Gallup [7] and Handel [8] have systematically sorted out the influencing factors of movie box office such as actors, marketing, story, and evaluation to predict box office revenue. Subsequent scholars carried out in-depth research under this research framework. Generally speaking, the box office success of a movie is mainly based on three dimensions: the characteristics of the movie (such as director, star, screenwriter, and genre), the strength of the marketing strategy (mainly through advertising budget, number of screens, trailers, etc.), and reviews (from critics and movie audiences, etc.). Study the influencing factors of movie box office from both the supply side and the demand side of movie products. Researchers explored many potential influencing factors, including film origin, film cost, schedule, director's influence, award-winning influence, professional rating, word of mouth, genre, celebrity influence, film content, reviews, cultural familiarity, and consumer factors. Among them, the three major factors of celebrity influence, comment, and word of mouth have received extensive attention [9]. In addition, in view of the huge economic impact of film sequel products in the film industry, scholars have begun to study the impact of this factor [10]. With the explosive growth and development of digital technology, movie consumers can express their opinions or attitudes towards products across space and time. Therefore, in recent years, electronic word of mouth (eWOM) in the form of online reviews has increased exponentially [11]. Many researchers study the impact of eWOM indicators on box office performance. With the development of big data technology, more scholars use social media and digital marketing activities as influencing factors to predict box office [12]. On the whole, traditional box office forecasting studies use factors such as budget, actors, directors, producers, story locations, screenwriters, screening time, music, screening locations, target audiences, and sequels as variables. Research based on the background of digital transformation extends the influencing factors to include social media topics, search engines, marketing activities, and other variables with connotative characteristics of digital consumption.

*2.2. Box Office Prediction Model Research.* Early box office prediction methods were based on audience surveys. Since Litman et al. (1989) put forward a model that affects movie box office income factors and movie rental income through regression analysis [13], the research on movie box office



prediction model methods has continued to advance. Scott Sochay (1994) made improvements based on the above model [14]. Representative scholars such as de Vany and Walls (2004) used the OLS model, and Deuchert et al. (2005) proposed a two-stage model [15]. Other researchers have carried out extensive linear regression research on this basis. Ramesh et al. (2006) first proposed a box office prediction model using neural network methods, which opened up the research on innovative methods of box office prediction models in the digital age [16]. Based on big data and machine learning technology, the accuracy of the box office prediction model has been further improved. Choudhery et al. (2017) constructed a polynomial regression model for box office prediction by extracting chat data to analyze user emotions and other three methods [17]. Although the accuracy of the neural network model is improved compared with the previous two prediction models, the results are still not satisfactory.

In summary, there is a solid research foundation on the influencing factors of movie consumption and box office forecasting models, and the framework of the influencing factor evaluation system including the main creative team, movie characteristics, marketing promotion, and word-of-mouth comments has been basically formed. And in terms of research methods, the research methods of statistical measurement models such as market survey questionnaire interviews and linear regression have gradually expanded to neural networks, machine learning, and data fusion in the context of big data. However, in previous studies, different research methods have only considered the linear effects of some factors on box office forecasts, and empirical research on box office forecasting models using machine learning and model fusion on the basis of fully considering the complex and comprehensive digital age influence factors is relatively lacking. This has laid a certain theoretical foundation for this research from influencing factors to the improvement of research methods.

### 3. Design of Characteristic Index System for Influencing Factors of Movie Box Office in the Era of Digital Economy

Based on the mature experience of film product attribute feature selection at home and abroad, combined with consumers' personalized characteristics and aesthetic preferences, this study focuses on the impact of digital environment elements on consumption logic under the background of digital transformation; explores the three-dimensional feature factors of consumers, film products, and digital environment, which have a great impact on film consumption in the digital era; and constructs an index system. In order to ensure the comprehensiveness of the evaluation of the characteristics of the influencing factors of film product consumption in the digital age, first of all, according to the personal influencing factors of movie consumption generally mentioned in the existing literature, the indicators of gender, age, education level, active area, and preference type are selected to reflect the basic information of the per-

sonal characteristics of movie consumption, aesthetics and preferences, and the influence of herd atmosphere. Second, fully consider the determinants of the film's main creative team and the characteristics of the film product. The cultural awareness of the core creative subject directors, screenwriters, and main actors, such as word-of-mouth popularity, box office appeal, number of movies, release schedule, 3D, and IMAX factors, is added to the film product feature evaluation indicators. In this way, the original value, artistic value, experience and emotional symbolic value, and cultural recognition related to the characteristics of the film product are measured. Third, it focuses on the most important changes in film consumption under the influence of the digital economy era, such as online social support, social marketing activities, and digital opinion leaders. Include the marketing activities in the environmental characteristics of the digital age, the popularity of public opinion, the number of publicity placements under the influence of Internet word of mouth, the platform, the amount of broadcast, the public opinion evaluation and popularity of online media, the score of Internet word of mouth, and the schedule and other factors. Based on the above reasons and the availability of data, the evaluation index system for the influencing factors of film consumption characteristics in the digital era set in this study, that is, the follow-up characteristic data collection system setting, is shown in Table 1.

### 4. Machine Learning Fusion Forecasting Model Construction and Demonstration

*4.1. Data Collection and Processing.* The data sources for this research mainly come from the professional database of Maoyan, a well-known professional film website in China, Sina Weibo, IMDB professional database, and WeChat official account platform. The relevant professional database mainly provides timely, accurate, and professional film creation and box office data analysis for practitioners in the domestic and foreign film industry. Among them, the Maoyan database has fully opened up the online movie information database, which is more suitable for studying the influencing factors of domestic movie consumption. Sina Weibo and WeChat are mainly used as the source of digital environment feature collection. In order to fully reflect the impact of environmental changes in the digital economy era on film consumption characteristics, considering the comprehensiveness and continuity of the data, the sample collection interval is the index information related to the consumption characteristics of all domestic films from 2017 to 2019. Preliminary data collection uses Python to complete the data capture and analysis. First, collect information about the personal characteristics of each movie consumer displayed on the website, and secondly, collect the culture, experience, and cognition information of the movie, such as the main creators and the company's topical discussion on social media, historical box office, number of representative works and movie awards, IP information, genres, and sequels. In addition, collect information on external environmental elements such as related distribution and promotion and the representative work of a film distribution

TABLE 1: Setting of the characteristic index system of influencing factors of movie box office in the digital age.

First-level indicator	Secondary index	Three-level indicators	Explanation of related indicators
Personal consumption characteristics	Basic information Aesthetics and preferences Herd mentality	Gender	Consumer gender distribution information
		Age	Consumer age distribution information
		Education level	Consumer education level distribution information
		Active area	Consumer active area distribution information
		Preferred movie type	Consumers' favorite movie genres in the past
		Film awards or nominations	Refers to all valid awards won by the film at major film festivals
		Movie type	The categories, types, or forms of films formed due to different themes or techniques, including 13 categories such as action, science fiction, and comedy
		Visual effect	Whether it belongs to 3D, IMAX, or giant screen
		Whether adapted from IP	Whether the movie is adapted from classic classics, best-selling novels, animation works, game works, etc.
		Whether the sequel	Does the movie belong to the sequel of a certain series of movies
Features of movie products	Core cultural value Experience and emotional value Cultural awareness	Director's box office appeal	Director's cumulative historical box office
		Starring box office appeal	The top ten leading actors in their respective cumulative historical box office
		Screenwriter box office appeal	The top three screenwriters have their cumulative historical box office
		Director topic discussion volume	Number of director's online topic discussions
		Leading topic discussion volume	Top ten leading actors in their respective online topic discussions
		The volume of screenwriting topics	The top three screenwriters discuss their respective online topics
		Number of masterpieces produced by the company	Cumulative number of masterpieces from major production companies
		Number of masterpieces produced by the company	Cumulative quantity of all masterpieces of major production companies
		Number of representative works issued by the company	Cumulative number of masterpieces of major issuing companies
		Trailer run times	Trailer network runs
Digital environment characteristics	Marketing activities Popularity of public opinion Internet word of mouth	Total number of trailers played	Cumulative play volume of trailer
		Distribution platform	The distribution of trailer delivery platforms
		Cumulative number of popular Weibo	Number of related hot Weibo discussions
		Cumulative Weibo interactions	Number of Weibo interactions
		Weibo topic discussion volume	Number of topics discussed on Weibo
		Cumulative number of official account articles	Number of related public account articles
		Cumulative article reading	Cumulative reading volume of related public account articles
		Cat eye score	Word-of-mouth score of Maoyan movie website
		IMDB score	IMDB word-of-mouth score
		Screening time/schedule	Film first round screening schedule

company to identify the company's ability, as well as the number of promotional materials, quantity, platform, topic level of professional mass social media, public opinion popularity indicators, and the impact of the planned cycle of

movie schedules. Subsequently, manual screening of units and text errors, as well as data cleaning of erroneous data, redundant data, and missing data due to the data transmission process, was carried out, totaling all the index

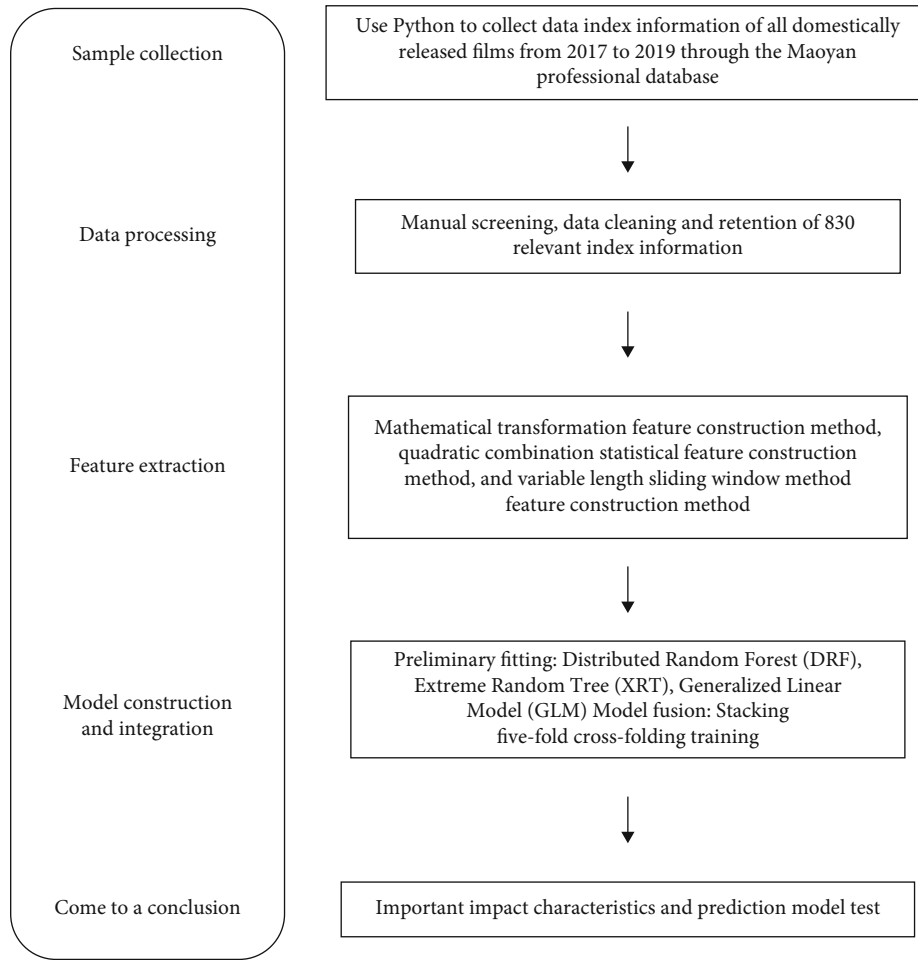


FIGURE 1: The research flow chart of machine learning and model fusion based on the analysis of the influencing factors of film consumption in the digital age.

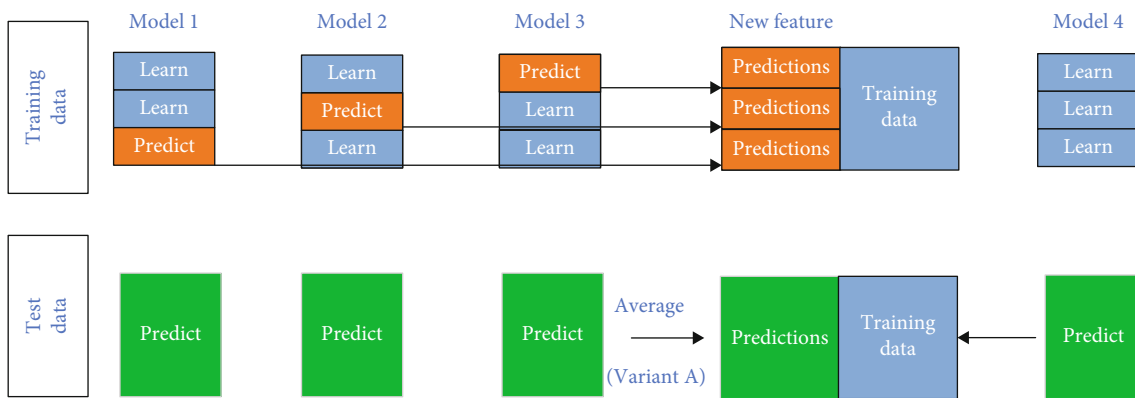


FIGURE 2: Stacking process model structure diagram.

information of 830 movies. In the future, new feature construction will be carried out according to research needs and specific scenarios. In view of the different data types having their own characteristics, different processing methods will be used to fit the research model.

4.2. *Research Method Selection.* The use of machine learning methods for movie box office prediction has made some research results in recent years, but most of the research only converts box office prediction from a regression problem to a classification problem. However, the use of classification

TABLE 2: Fitting effect analysis of distributed random tree forest model.

DRF: distributed random forest	
Model	DRF_1_AutoML_20200707_105158
Frame	automl_training_Key_Frame__movie_r2.hex
Description	Metrics reported on out-of-bag training samples
Model_category	Regression
MSE	0.039718
RMSE	0.199295
r2	0.941232
Mean_residual_deviance	0.039718
mae	0.134591
rmsle	0.023422

methods to predict the box office will lose a lot of characteristic information, which may cause certain restrictions on the use of prediction results. Feature engineering methods can extract core features, which have a vital impact on the accuracy of prediction models [18]. Through the innovative integration of machine learning feature engineering and regression models that process multiple data, it is more conducive to accurately assess the influencing factors and box office expectations of movie consumption in the digital age. Therefore, this study first uses Python computer program design language-related data directed crawler (requests + bs4 + re) library to complete the analysis of digital movie product consumers' personal characteristics, product characteristics, and digital environment network interaction behavior characteristics. Through manual screening, data cleaning, and preprocessing, combined with feature engineering research methods in the field of machine learning, Scikit-learn is used for feature extraction and feature construction. Then, according to the diversity of data types related to movie influencing factors. The innovation uses the stacking method in the machine learning integration algorithm to fuse models based on the fivefold cross-folding training method for distributed random tree forests (distributed random forest), extremely randomized trees (extremely randomized trees), and generalized linear model which are good at handling different data types. It has a higher fit and model accuracy in feature mining and model construction, so as to more effectively grasp the relevant feature factors that affect movie consumption and more accurately predict the future movie box office.

**4.3. Research Idea Design.** This research is based on the exploratory structure and deep insight of data characteristics and innovatively adopts the model fusion perspective for machine learning applications. The design of the research ideas is shown in Figure 1. First of all, comprehensive preliminary research and literature research combining the design of influencing factor index system and accurate data collection are the basic guarantees for the construction of feature engineering models. Good data preprocessing can explore the direction and accuracy of model training. Second, carry out data cleaning and screening to retain valid information. After that, input the effective data reflecting

the different influencing factors of the movie into different feature learning models, extract the features of the corresponding movie, and try to construct new features. Due to the very large scales of different types of variables in this study, exploratory statistical analysis found that data such as cumulative box office, first week box office, and star cumulative box office are all exponentially distributed, so logarithmic transformation of these features can construct new features. Finally, the stacking model fusion method is selected to construct the box office prediction model, and the three basic models are learned and fused by designing the feature vector fivefold crossfolding training method to construct a more accurate prediction model. In this way, it can more accurately identify the feature vector that conforms to the film consumption in the digital age and reveal the source of film box office revenue.

The processing of high-dimensional and complex data is a difficult point in machine learning. In traditional classification algorithms, it is difficult to deal with problems in practical applications, presenting problems such as low accuracy and overfitting. Stacking model is essentially a hierarchical structure, which is good at dealing with model fusion problems, and it is also particularly suitable for model training and learning that deal with multidimensional complex factors. Through the fitting and learning of different types of models, fusion builds an innovative fusion model that is more in line with the characteristics of the data. It is very suitable for the complex and multivariate influencing characteristic variable types of this research and the actual needs of accurate box office prediction. Figure 2 shows the basic process structure of this method.

**4.4. Model Construction and Empirical Analysis.** The feature engineering construction method of machine learning is used to analyze, collect and construct features, and determine which consumption features are the most important, which plays a role in the performance of the predictive model. It helps to avoid errors in human factor judgments and some inertial problems of traditional statistical measurement models and helps to obtain a more explanatory characteristic variable system. According to the data characteristics of the influencing factor index system, the following three types of classic models are used for fitting, respectively,

TABLE 3: Fitting effect analysis of extreme random tree model.

ERT: extremely randomized trees	
Model	ERT_1_AutoML_20200707_105158
Frame	automl_training_Key_Frame__movie_r2.hex
Description	Metrics reported on out-of-bag training samples
Model_category	Regression
MSE	0.037427
RMSE	0.193462
r2	0.944621
Mean_residual_deviance	0.037427
mae	0.132382
rmsle	0.022773

TABLE 4: Analysis of fitting effect of generalized linear model.

GLM: generalized linear model	
Model	GLM_1_AutoML_20200707_105158
Frame	automl_training_Key_Frame__movie_r2.hex
Description	.
Model_category	Regression
MSE	0.051241
RMSE	0.226365
r2	0.924182
Mean_residual_deviance	0.051241
mae	0.16952
rmsle	0.027431

TABLE 5: Analysis of fitting effect of fivefold crossfolding training fusion with three models.

Stacked ensemble	
Model	StackedEnsemble_AllModels_AutoML_20200707_105158
Frame	automl_training_Key_Frame__movie_r2.hex
Description	.
Model_category	Regression
MSE	0.005488
RMSE	0.074081
r2	0.99188
Mean_residual_deviance	0.005488
mae	0.050961
rmsle	0.008718

and the stacking model fusion method is used to perform fivefold crossfolding training on different models to construct a new fusion model. This makes the fusion model stronger in fusion and generalization and forms a model structure that is more suitable for the identification of influencing factors of film consumption and box office prediction in the digital age.

#### 4.4.1. Distributed Random Tree Forecast Model Experiment.

Bernard et al. proposed that stochastic forest is one of the

most classical data processing models of integrated learning algorithm. It provides users with reasonable and effective classification label information by using the integrated thought, thus providing reliable and effective data information recommendation [19]. Fernández-Delgado et al. found that the random forest algorithm has the best classification performance by comparing the classification performance of 179 classification algorithms [20]. Lizhi et al. found that the distributed random forest algorithm in Spark is more suitable for feature learning of two-dimensional variables

TABLE 6: Analysis of fusion combination factor for fivefold crossfolding training with three models.

Names	Coefficients	Standardized_ coefficients
Intercept	-0.1308	7.6077
XRT_1_AutoML_20200707_195606	0.9967	0.7847
DRF_1_AutoML_20200707_195606	0.0043	0.0033
GLM_1_AutoML_20200707_195606	0.0166	0.0129

[21]. The data collection conforms to the data structure and characteristics of movie consumption factors in the digital age. The empirical study also shows a good fit effect. Table 2 shows that the goodness of fit with this model reaches about 94.12%, and the prediction error RMSE of the model reaches 19.9%.

#### 4.4.2. Extreme Random Tree Prediction Model Experiments.

The extreme random tree algorithm proposed by Geurts et al. [22] is very similar to the random forest algorithm, but the extreme random tree features are randomly selected. Selecting the best partitioning feature with the specified threshold as the optimal partitioning attribute not only guarantees the utilization of training samples but also reduces the final prediction bias, so it is superior to the results obtained by random forests to some extent. Therefore, it is also used as a prediction model method to carry out experiments. The results obtained in this paper also conform to a high level of goodness of fit, basically reaching about 94.46%, and the RMSE prediction error reaches 19.3%, as shown in Table 3.

#### 4.4.3. Experiments of Generalized Linear Prediction Model.

The generalized linear model is an extension of the general linear model. It establishes the relationship between the expected value of the response variable and the predicted variable of the linear combination through a join function. It is characterized by not forcibly changing the natural measures of the data. The data may have a nonlinear and non-constant variance structure, or it may be the most popular machine learning algorithm at present. This study also uses this algorithm to fit according to the structure characteristics of the data indicators. The results of the analysis are relatively consistent with the data characteristics, reaching a good fit of 92.41%, but the RMSE prediction error is as high as 22.63% (see Table 4).

#### 4.4.4. Triple Model Fusion Fivefold Crosstraining Experiment.

Fitting the data of movie consumption characteristics with the above three models, we can find that, first, the selection of the initial index system is more effective, making these basic characteristics representing movie consumption more regular. At the same time, the three algorithms have more than 90% fitting accuracy and strong explanatory power, but there is still room for further improvement of prediction

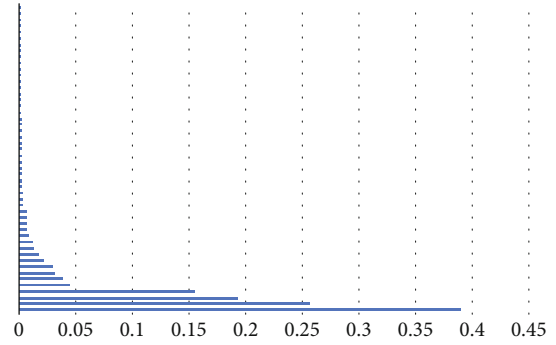


FIGURE 3: Feature extraction of influencing factors of movie consumption characteristics in the digital age.

accuracy. In order to further explore the consumption characteristics, the stacking model fusion method was used to train the three models with fivefold crossfolding, and a more accurate model was obtained. The goodness of fit reached 99.18%, while the RMSE was 7.4%, and the RMSLE was significantly lower than the classification prediction error of the first three classical models which was only 0.8%. This fully demonstrates that the features extracted by learning from this model are very consistent with the characteristics and actual results of the database. The model has more generalization ability and basically matches the feature structure of the current factors affecting movie consumption perfectly (see Table 5).

The results of this fusion model are integrated by the three model algorithms mentioned above, and their combination coefficients are shown in Table 6.

4.5. Results and Discussion. Based on the above model learning results, it can be found that through innovative model fusion training, the goodness of fit is higher and the prediction bias is lower than that of a single prediction model. In the digital economy era, the extraction of influencing factors of movie consumption is more accurate and can provide more effective box office prediction model scheme. Based on the analysis results, this study further discusses and analyzes which extracting features can better reflect the explanatory and influence of movie consumption in the digital age. Feature extraction and learning are carried out through different models. Important indicators in the influence feature variables of digital movie consumption are shown in Figure 3.

Therefore, it can be found that the most influential feature is the cumulative historical box office of star writers in the main body of the core content creator, which fully reflects the importance of the current market on the core content. First, the writer is the core creator of the current creative source of digital content products, and also the source of IP core stories. Past box office represents the creative ability of the writer, the cultural and artistic value of the work and the docking ability of the market, highlighting the significance of content as king.

Second, digital marketing has become an important influencing factor of movie consumption. Essential changes have taken place in the form of movie marketing promotion

in the digital age. The broadcast of marketing materials put on the Internet has become an important feature affecting movie consumption. Movie consumption in the digital age has a wider audience. In the market of digital content products, the voice and influence of social media play an important role. Precise delivery and distribution mechanism based on Internet platform can help achieve digital marketing effect. Thirdly, the cumulative historical box office of star creators shows that the past artistic performance and recognition of star creators are very important, and star is still the core value creator of content products. Fourth, the type of movie still has an important impact. Although this factor has been proven by many studies to be closely related to movie consumption, special types such as love, action, and science fiction still become an important factor to trigger the resonance of movie consumption and stimulate market vitality. Fifth, hot spots of public opinion have become important variables affecting movie consumption, including different types of self-media comments and word-of-mouth communication and discussion, such as Weixin Public Number and Weibo topic discussion.

## 5. Conclusions

Based on the empirical results of these factors affecting movie consumption, the following suggestions are given on how to improve the consumption related to digital content in the future:

First of all, attach great importance and increase capital investment to creative subjects of high-quality cultural content and treat the flow effect with caution. Along with the continuous innovation of digital content form, it brings consumers a more pleasant consumption experience but also changes people's traditional consumption habits and consumption concepts. Secondly, further standardize the network environment and strengthen the network ecological governance. The most important influencing factors of consumption of digital content products are the guidance and evaluation of network public opinion. The network environment should further be standardized; major movie and television websites should do a good job in related management, focus on "zombie" number and accounts with malicious scoring records, and rectify the black industry chain that the network breeds. Third, encourage the construction of a diversified digital content value evaluation system. For cultural creative consumers, big data on the Internet only means the display and prediction of large probability events, which can only be used as reference. Digital content products are essentially cultural creative products. Its cultural value and aesthetic experience cannot be pale and shallow only represented by a series of data. Finally, encourage content providers such as digital content creative subject, production producer, and dissemination subject to adhere to the original intention of content creation. Make good use of digital diffusion channels and create a win-win situation between content providers and consumers by using "big data." However, the prediction model used in this article is sensitive to noise, and the prediction accuracy needs to be

further improved. These two shortcomings are also the direction for future work.

## Data Availability

The data used to support the findings of this study are included within the article.

## Conflicts of Interest

All the authors do not have any possible conflicts of interest.

## Acknowledgments

This work is supported by the National Natural Science Foundation of China (Grant No. 71704102).

## References

- [1] M. Zhao, A. Jha, Q. Liu et al., "Faster mean-shift: GPU-accelerated clustering for cosine embedding-based cell segmentation and tracking," *Medical Image Analysis*, vol. 71, article 102048, 2021.
- [2] Z. Chu, M. Hu, and X. Chen, "Robotic grasp detection using a novel two-stage approach," *ASP Transactions on Internet of Things*, vol. 1, no. 1, pp. 19–29, 2021.
- [3] M. Zhao, Q. Liu, A. Jha et al., "VoxelEmbed: 3D instance segmentation and tracking with voxel embedding based deep learning," 2021, <https://arxiv.org/abs/2106.11480>.
- [4] W. Wang and D. Bin, "Machine green light system' and' algorithmic matrix movie' - the impact of artificial intelligence on the film production industry," *Contemporary Movies*, vol. 12, pp. 30–36, 2020.
- [5] T. Siegel and W. Bros, *Signs Deal for AI-Driven Film Management System*, no. 1, 2020Hollywood Reporter, 2020.
- [6] R. Handel et al., *How Hollywood Understands Audiences*, Chinese Press, 2014.
- [7] B. Ayoub, "George Gallup in Hollywood," *Pacific Historical Review*, vol. 77, no. 4, pp. 693–695, 2008.
- [8] L. A. Handel, *Hollywood Looks at its Audience. A Report of Film Audience Research*, The University of Illinois Press, Urbana, IL, 1950.
- [9] F. Peng, L. Kang, S. Anwar, and X. Li, "Star power and box office revenues: evidence from China," *Journal of Cultural Economics*, vol. 43, no. 2, pp. 247–278, 2019.
- [10] B. Belvaux and R. Mencarelli, "Prevision model and empirical test of box office results for sequels," *Journal of Business Research*, vol. 130, no. 1, pp. 38–48, 2021.
- [11] H. Ma, J. M. Kim, and E. Lee, "Analyzing dynamic review manipulation and its impact on movie box office revenue," *Electronic Commerce Research and Applications*, vol. 35, article 100840, 2019.
- [12] Z. Wang, J. Zhang, S. Ji, C. Meng, T. Li, and Y. Zheng, "Predicting and ranking box office revenue of movies based on big data," *Information Fusion*, vol. 60, pp. 25–40, 2020.
- [13] B. R. Litman and L. S. Kohl, "Predicting financial success of motion pictures: The '80s experience," *Journal of Media Economics*, no. 2, pp. 35–50, 1989.
- [14] S. Sochay, "Predicting the performance of motion pictures," *Journal of Media Economics*, vol. 7, no. 4, pp. 1–20.

- [15] A. de Vany and W. D. Walls, "Uncertainty in the movie industry: does star power reduce the terror of the box office?," *Journal of Cultural Economics*, vol. 23, no. 4, pp. 285–318, 1999.
- [16] S. Ramesh and D. Delen, "Predicting box-office success of motion pictures with neural networks," *Expert Systems with Applications*, vol. 30, no. 2, pp. 243–254, 2006.
- [17] D. Choudhery and C. K. Leung, "Social Media Mining: Prediction of Box Office Revenue," in *Proceedings of the 21st International Database Engineering & Applications Symposium on - IDEAS 2017*, 2017.
- [18] Y. Ru, B. Li, J. Chai, and J. Liu, "A movie box office prediction model based on in-depth learning," *Journal of China Media University: Natural Science Edition*, vol. 26, no. 1, pp. 30–35, 2019.
- [19] S. Bernard, S. Adam, and L. Heutte, "Dynamic random forests," *Pattern Recognition Letters*, vol. 33, no. 12, pp. 1580–1586, 2012.
- [20] M. Fernández-Delgado, E. Cernadas, S. Barro, and D. Amorim, "Do we need hundreds of classifiers to solve real world classification problems?," *Journal of Machine Learning Research*, vol. 15, pp. 3133–3181, 2014.
- [21] M. Lizhi, D. Jiyao, and L. Chong, "Breast cancer risk prediction analysis based on Spark and random forest," *Computer Technology and Development*, vol. 29, no. 8, pp. 142–146, 2019.
- [22] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Machine Learning*, vol. 63, no. 1, pp. 3–42, 2006.



## Research Article

# Big Data and Deep Learning-Based Video Classification Model for Sports

Lin Wang,<sup>1</sup> Haiyan Zhang ,<sup>2</sup> and Guoliang Yuan<sup>2</sup>

<sup>1</sup>Department of Physical Education, North China University of Science and Technology, 063210 Tangshan, Hebei, China

<sup>2</sup>College of Physical Education, Hengshui University, Hengshui, 053000 Hebei, China

Correspondence should be addressed to Haiyan Zhang; zhy18903182110@126.com

Received 28 August 2021; Revised 8 September 2021; Accepted 14 September 2021; Published 7 October 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Lin Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Information technologies such as deep learning, big data, cloud computing, and the Internet of Things provide key technical tools to drive the rapid development of integrated manufacturing. In recent years, breakthroughs have been made in big data analysis using deep learning. The research on the sports video high-precision classification model in this paper, more specifically, is the automatic understanding of human movements in free gymnastics videos. This paper will combine knowledge related to big data-based computer vision and deep learning to achieve intelligent labeling and representation of specific human movements present in video sequences. This paper mainly implements an automatic narrative based on long- and short-term memory networks to achieve the classification of sports videos. In the classical video description model S2VT, long- and short-term memory networks are used to learn the mapping relationship between word sequences and video frame sequences. In this paper, we introduce an attention mechanism to highlight the importance of keyframes that determine freestyle gymnastic movements. In this paper, a dataset of freestyle gymnastics breakdown movements for professional events is built. Experiments are conducted on the data and the self-constructed dataset, and the planned sampling method is applied to eliminate the differences between the training decoder and the prediction decoder. The experimental results show that the improved method in this paper can improve the accuracy of sports video classification. The video classification model based on big data and deep learning is to provide users with a better user experience and improve the accuracy of video classification. Also, in the experiments of this paper, the effect of extracting features for the classification of different lifting sports models is compared, and the effect of feature extraction network on the automatic description of free gymnastic movements is analyzed.

## 1. Introduction

With the rapid development of computers, networks, multimedia, and other related technologies, multimedia data has shown an exponential growth trend. A video is a common form of multimedia data, and it is one of the important components of multimedia data [1], which is closely related to our daily life. Video contains the richest data information, with a more complex structure and a large amount of data. Faced with such a huge video data, automatic video description can better manage and utilize these rich video resources, which can help users to improve the indexing speed as well as the search quality of online videos, so that they can play a greater role. For people with impaired vision, the automatic description of videos and combined with text-to-

speech conversion technology converts the text within the computer into continuous natural language for communication. It can help them to understand the content in the video better, thus making life easier for the visually impaired. In the field of automatic video description research, automatic human action-based video analysis and understanding have gradually become a popular research problem in computer vision and pattern recognition in recent years. It has a wide application prospect in the fields of intelligent life assistance, advanced human-computer interaction, and content-based video retrieval and is closely followed by researchers at home and abroad [2].

Existing research results for high-precision classification algorithms and their conceptual drift still rely mainly on data structures, and algorithm optimization on data mining

as well as detection of conceptual drift is still mainly done by standalone computers with limited computational resources.

The growing and escalating levels of data and data complexity make it insufficient to rely solely on the algorithm itself and single-computer computing resources [3]. The use of distributed computing platforms to cope with the huge consumption of time complexity and space complexity of algorithms in big data environments and to address the problem of conceptual drift in data streams has become a major concern.

Faced with the current problems in the research of sports video high-precision classification models [4], such as low-level video features cannot accurately reflect high-level human semantic concepts, high time complexity, and low recognition accuracy of action recognition algorithms in traditional RGB videos, and the use of single features cannot meet the massive growth of existing video data and its recognition of complex actions, the study of the automatic description of videos represented by competitive sports events has important theoretical research significance and extensive practical application value. In terms of theoretical research, the study of the automatic description of sports video is a cross-cutting topic that integrates multiple disciplines such as big data analysis, machine learning, pattern recognition, video analysis, computer vision, and cognitive science, which provides a good research object for these fields, and its in-depth research can promote the development of related disciplines. With the progress of deep neural network research and the emergence of large-scale datasets in the fields of image classification and object recognition, a number of approaches have attempted to use convolutional neural networks to learn the semantic representation of images and then use recurrent neural networks to achieve their correspondence with natural language. Traditional supervised learning is mainly single-label learning, while real-life target samples are often complex, have multiple semantics, and contain multiple labels [5].

In recent years, automatic video understanding has gradually become a popular research direction in the field of computer vision [6–8]. Compared with image content research, the content of video contains more information, and a single label is not able to completely characterize the content of the video, so most of the problems for automatic video content understanding are multilabel problems. A number of learning algorithms on multilabeling have been proposed in existing research, and based on the problem-solving perspective, these algorithms can be divided into two major categories: the first category is based on problem transformation approaches, where the main difficulty of multilabeling learning lies in the explosive growth of the output space, and to cope with the exponential complexity of the label space, the correlation between labels needs to be mined. Effective mining of correlations between labels is the key to the success of multilabel learning. According to the strength of mining correlations, multilabel algorithms can be divided into three order strategies. First-order strategy: ignore the correlations between labels, e.g., decompose multilabel into multiple independent binary classification problems. Second-order strategy: consider pairwise correla-

tions between labels, such as ranking relevant and irrelevant labels. Higher-order strategies: consider correlations between multiple labels [9], such as considering the effects of all other labels for each label. The second category is based on algorithmically applicable methods. Problem transformation-based methods focus on transforming problem data to make it applicable to existing algorithms; algorithm-applicable methods are those that extend for a particular algorithm to be able to handle multilabel data, improve the algorithm, and apply the data. The video classification model uses a 3D convolution kernel to process space and time dimensions at the same time. However, the 3D convolution model is shallow and has a huge amount of parameters, which is very bloated. Finally, C3D is used. This model achieves the same as the 2014 dual-stream method, accuracy of close video behavior classification. It uses 3D convolution and 3D pooling and fully connected layers to form an 11-layer shallow network. Its biggest advantage lies in speed. However, the size of the C3D model reaches 321 MB, which is even larger than 152-layer ResNet 235 MB model. Such a model is difficult to train and cannot be pretrained on a large-scale image data set like ImageNet. The shallow network also limits the classification performance of the model.

## 2. Related Work

Early approaches to automatic video description were rule-based. A language model is used as the basis for predicting the subject, predicate, and object and then complementing the final description of the other constituent videos. For example, the literature describes human activities by introducing a behavioral concept hierarchy, and the literature uses a semantic hierarchy to learn semantic relations between different segments. The literature uses conditional random fields to model objects and activities and generates semantic features for description. In addition, the literature proposes a unified framework consisting of a semantic language model, a deep video model, and a joint embedding model to learn associations between videos and natural sentences. However, all of the above approaches rely excessively on well-defined rules and are limited by fixed syntactic structures, resulting in generated sentences that are too rigid for everyday descriptions. With the advances in deep neural network research and the availability of large-scale datasets in the fields of image classification and object recognition, a number of approaches have attempted to use convolutional neural networks to learn the semantic representation of images and then use recurrent neural networks to implement their correspondence with natural language.

The literature proposes a long short-term memory neural network, which effectively solves the problem of artificially prolonged time tasks that are difficult to solve by RNNs, and addresses the problem that RNNs are prone to gradient disappearance. The literature introduces the forgetting gate mechanism, which enables the LSTM (long short-term memory) to reset the state. The literature proposes a bidirectional long short-term memory neural network (BLSTM), which is one of the most widely used LSTM models. LSTM is a member of deep learning techniques,

and its basic structure is complex and has high computational complexity, which makes it more difficult to perform deeper learning, for example, Google Translate also only applies 7-8 layers of LSTM network structure. The literature is the first implementation of direct text generation from video, which uses convolutional neural networks to extract features from all frames in the video and then perform average pooling before sending them to the LSTM to decode and generate text. The S2VT proposed in the literature uses a long- and short-term memory network in both the encoder and decoder. The literature uses different models trained on different kinds of features thus to generate the description of the video and then uses an evaluation network to evaluate the correlation between the generated sentences and the video features and selects the best correlation as the final video description. The literature uses multiple types of features such as image features, video features, ambient sound features, speech features [10], and kind features to fuse them as a representation of the video. The literature proposes a transfer unit to model the high-level semantic properties of incoming images and videos that are presented as supplementary knowledge for video representations to facilitate sentence generation.

In the methods of sports video content analysis and recognition research, most of the existing sports video research focuses on the semantic analysis of sports and its recognition research. And in the study of sports classification breakdown, except for ball sports such as football, basketball, golf, and badminton, other sports research is rarely involved. The literature borrows a dynamic Bayesian network to analyze sports video, then extracts image features by Kalman filtering principle and uses EM algorithm to complete DBN parameter learning. The literature proposes a semantic content analysis model based on perceptual concepts and finite state machines to automatically describe and detect meaningful semantic content in sports videos. The literature proposes a pose constraint-free upper-body human target detection algorithm. The literature proposes two novel nonlinear feature fusion models and designs an automatic sports video classification algorithm based on the novel features and support vector machines. The literature proposes an automatic detection method based on convolutional neural networks for the detection problem of multiscale athletes in sports videos [11]. The literature uses feature filtering and support vector machines for sports video recognition, which improves the accuracy of sports video recognition and speeds up the speed of sports video recognition. There are offline computing, real-time stream computing, and two types of computing frameworks.

### 3. Video Classification Model Based on Big Data and Deep Learning

Big data and the deep market is entering a period of rapid development with scale applications in industries such as healthcare. In the market, there are video classification modes based on partial differential equations, and video classification modes based on big data and deep learning, education, the Internet, and commerce. The huge amount

of data resources such as user's behavioral characteristics need to be further analyzed and mined to build commercial sports video classification models to provide to producers and sellers to improve the product experience. Big data and deep learning technologies will not only help to increase the added value of the product but also maximize the value of the customer experience. Countries now have a huge market for big data technology and also have the data resources and technology accumulation base for big data to enable more accurate sports video classification [12].

#### 3.1. Streaming Data Model for Sports Video Big Data

3.1.1. *Basic Introduction to Streaming Data for Sports Video Big Data.* Streaming data is a set of sequential, large, fast, and continuous data sequences. In general, a data stream can be regarded as a dynamic data collection that grows indefinitely over time. It is used in the fields of network monitoring, sensor network, aerospace, meteorological measurement and control, and financial services. The study of streaming data is the study of a new type of data processing model and therefore requires a different approach to data query mining than the traditional method, adapted for streaming data application scenarios, to implement a streaming data query and mining algorithm based on a distributed streaming computing framework and form a prototype algorithm system to query and mine the latest data as fast as possible and give results [13], solving the streaming data loop practical problems in the environment. Whether it is in the field of data mining or database research, more and more experts and scholars at home and abroad are paying attention to the research to stream data will be a focus that cannot be ignored in the future, and top conference sessions such as SIGMOD, ICDE, VLDB, and ICDM will have related topics every year. The research on data analysis and processing for big data is mainly in two aspects, stream data query and stream data mining. Data flow management system is mainly for the research of data flow query and on this basis continue to deepen and build some data management solutions for specific application background items; stream data analysis and processing methods have significant differences with the traditional data processing; first of all, stream data is real-time dynamic data flow and is dynamic relative to the traditional static data form. Secondly, the data elements of streaming data are independent and random, and the dynamic changes of data flow are not easy to predict. More importantly, the real-time nature of streaming data and the large volume of data makes it more time and space complex for the system compared to traditional data. Streaming data requires real-time in-memory processing and outputting the results, rather than traditional data by storing it to disk first and then computing it. Streaming data processing differs significantly from traditional data processing in terms of data form, processing characteristics, data storage methods, update rates, and real-time requirements (as shown in Table 1).

The first thing to consider with stream data mining algorithms is what data to process. The most accurate solution is, of course, the one that is most accurate for all.

TABLE 1: Characteristics of stream data processing and comparison of traditional requirements with traditional data.

Feature requirements	Traditional data	Streaming data
Data format	Continuous and stable data flow	High-speed data flow
Data storage method	Passive	Initiative
Efficient	Low	High

All the data that have arrived are mined without any omission, but the characteristics of stream data make the method impossible to achieve. Stream data processing model is the study of data stream summary model, that is, the analysis of which part of the data stream needs to be processed for analysis. For stream data, its large data volume characteristics is so often taken according to the algorithm needs to select a range of data within the accuracy requirements for processing, rather than all data as the processing object [14]. The current data stream processing is mainly based on the approximation of the data selection time range, so the selection of the range can be divided into snapshot model, boundary marker model, and sliding window model.

### 3.1.2. Stream Data Mining Algorithm Features

- (1) Incrementalization: traditional data mining algorithms are mined for offline datasets, where data may be accessed by the algorithm multiple times to get the final converged model, whereas real-time mining algorithms emphasize making full use of the latest data for predictive model training or finding knowledge and patterns in the latest data. Therefore, real-time mining algorithms need to be able to continuously update the original training model using the latest incremental data
- (2) Anti-interference: traditional data mining algorithms run data mining with a single background principle or pattern of the target data, i.e., the target data set is influenced by the same factors or contains knowledge of the same pattern, whereas the actual existence of real-life stream data contains a model that may change and may have multiple variations in a form called “conceptual drift” [15]. Therefore, real-time mining algorithms need to be able to effectively resist the interference of the concept drift phenomenon; otherwise, the extracted patterns or the trained models will be outdated and inaccurate
- (3) Labeling problem: for data mining algorithms, classification, and regression in supervised learning, the dependent variables of their training sets need to be labeled. For traditional data mining algorithms, although the cost of training set labeling is high and time-consuming, it can still accomplish the set goal because it is offline mining; for real-time data

mining, because the training data flows into the system incrementally and in real time, the timing of labeling is fleeting, and further research on semisupervised algorithms is needed to form a closed-loop process for training set labeling

**3.2. Big Data Distributed Computing Framework.** The video big data streaming computing architecture requires several requirements such as low latency, scalability, high reliability, and fast recovery. The computing platform needs to be able to track the execution of each message and deliver the message quickly, and the forwarding delay needs to be below milliseconds; the computing platform needs to support horizontal scaling so that when the computing needs require system expansion [16], it is easy to operate and does not affect the existing data processing operations; the computing platform needs to be highly stable, support the rapid detection of node failure, and be able to resist the avalanche phenomenon.

**3.2.1. Storm.** The storm is an open-source real-time distributed computing system developed by Twitter, which features scalability, high system fault tolerance, etc. Storm defines the original language for developers to use based on platform’s real-time computing and development and has convenient APIs for developers to use, such as real-time online queries, machine learning, and data mining. Storm can be used by configuring ZooKeeper, thereby facilitating the scaling of its distributed clusters. The computational principles are as follows.

$$F_t = \frac{1}{n} \sum_{i=1}^n X_i Y_i \cdot \left( \frac{x - \mu}{\sigma} \right). \quad (1)$$

Storm can process millions of messages in a second, in terms of sports video classification. The development languages for Storm are Clojure and Java, and non-JVM languages can communicate with Storm via stdin/stdout using the JSON protocol. One of the advantages of Storm is that its topology writing and message processing components can be freely used with a variety of programming languages. The most important feature of Storm is that it guarantees that messages are processed, a feature that will benefit the reliability of the entire architecture. By ensuring that messages are processed, Storm ensures that the entire streaming processing layer of the architecture is guaranteed to process messages. The computational framework is shown in Figure 1.

**3.2.2. YahooS4.** YahooS4 with distributed, pluggable, scalable, and partitioned fault-tolerant features was the first system used to process ad click traffic data. S4 uses the Actor architecture model using a simple programming interface to enable large-scale development. S4 has no central control node, and each working node is of equal status, which makes the system highly usable. S4 has the Actor computing the processing unit PEs in the S4 system pass the completed data through the Actor architecture in the form of events. Each PE is independent of each other and interacts with each

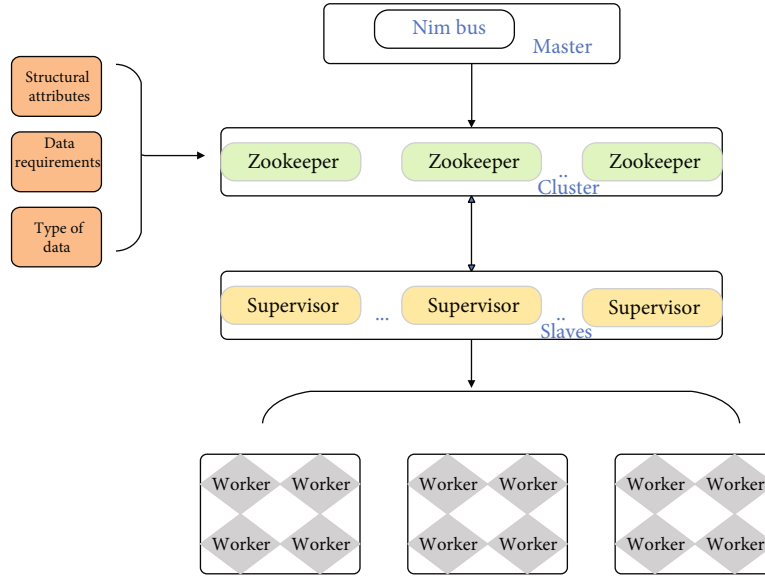


FIGURE 1: Storm stream data computation framework diagram.

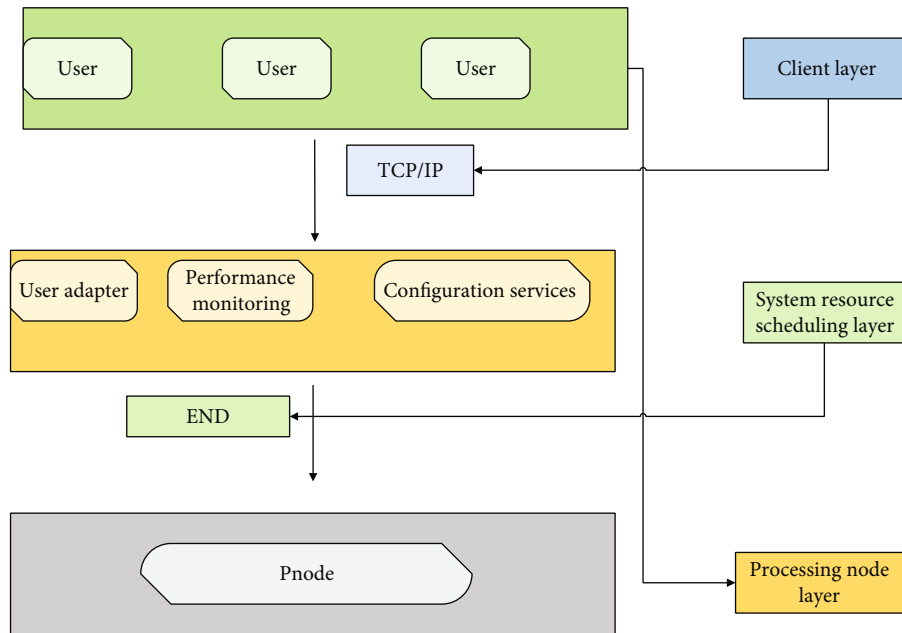


FIGURE 2: S4 system framework diagram.

other by issuing events and interaction events. The specific principle formula is

$$G(x, y) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{i=1}^n X_i Y_i \cdot \frac{1}{n} \sum_{i=1}^n X_i. \quad (2)$$

S4 draws on the MapReduce model and solves the single point of fault tolerance problem. The independent peer-to-peer architecture not only improves the scalability of the cluster but also ensures the efficiency of the system. S4 system supports Java language development and modular packaging. The functional components of S4 are divided into

three major categories: Clients, Adapters, and PNodeCluster. Figure 2 shows the framework of the S4 system.

3.3. *Deep Learning Based on Video Classification.* Deep learning is a recent area of research that has received much attention and plays an important role in machine learning. The history of deep learning development, in terms of timeline, can be seen as three stages of neural network development. The first generation of neural networks [17], which can be called artificial neural networks, was studied starting from the proposal of the M-P model. Figure 3 illustrates the M-P structural model.

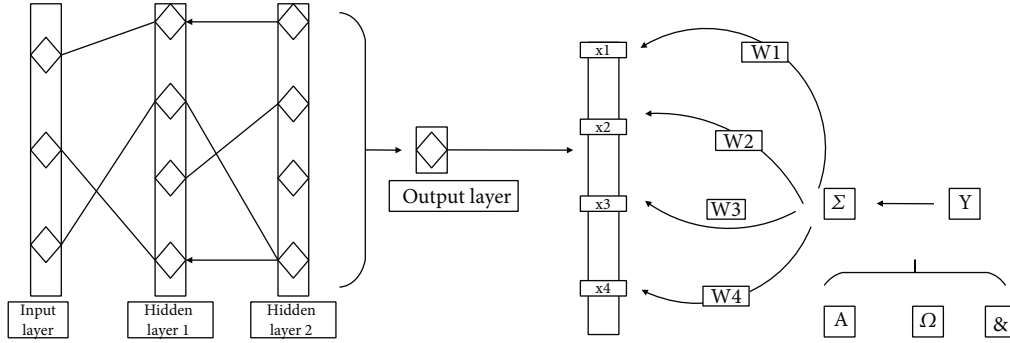


FIGURE 3: Basic model diagram of neural network.

The backpropagation algorithm for artificial neural networks was proposed, which not only gave hope to the development of machine learning but also opened the statistical model-based machine learning that is still of deep research significance today. The third stage is deep learning, and in the second stage, deep learning enters a period of rapid development by pretraining to quickly train deep belief nets. In the third phase, researchers reduced the top 5 error rate of the ImageNet image classification problem to 15% and deep learning entered an explosive period [18, 19].

Deep learning is a part of machine learning, which can be divided into two research phases, shallow machine learning and deep learning, in terms of research history. Shallow machine learning models and deep learning models have commonalities and important differences. Shallow machine learning models do not use distributed representations and require human extraction of manual features in feature extraction, which is not only time-consuming, heavy, and difficult but also requires accurate domain knowledge of the relevant specialty when researching a particular domain. The model itself can only make further prediction or classification based on the extracted features, and the quality of human-extracted features directly determines the performance of the whole system to a large extent. Deep learning can learn to get the essential features of the whole data set from a smaller number of data samples when performing feature extraction. A nonlinear deep network structure is obtained by training the network to learn and achieve a distributed representation of the input data by representing complex functions with a small number of parameters. The essence of deep learning is to improve the accuracy of classification or prediction by building machine learning models with a number of hidden layers and huge amounts of training data to learn the desired features. The hidden layer in deep learning is equivalent to a linear combination of input features, and the weight between the hidden layer and the input layer is equivalent to the weight of the input features in the linear combination [20], and the capability of the deep learning model grows exponentially with the depth of the network, and its specific principle is formulated as

$$St = \frac{1}{n} \sum_{i=1}^n X_i Y_i. \quad (3)$$

Recurrent neural networks (RNNs) are a special kind of neural network structure inspired by the fact that humans rely on experience and memory during cognition. The reason why RNNs are called recurrent neural networks is that RNNs assign not only a memory function to the input of the previous moment but also the input of the next moment is referred to the memory of the previous moment; the current output of a sequence is jointly determined by its input and the output of the previous sequence determines. The specific process is manifested in that the current output is computed by applying it to the output remembered from the previous moment. RNNs differ from CNNs in that in RNNs, the input data have a temporal order of precedence, thus forming a sequence. This is the key difference between RNNs and other neural networks and is the reason why the “loop” can be established [17]. The nodes between the hidden layers, which are unconnected in CNN, become connected in RNN, and the input of the hidden layer contains the output of the input layer and the output of the hidden layer at the previous moment. Its mechanism is illustrated in Figure 4. The hidden layer of the simplest structure of the RNN is expanded in time.

*3.4. Algorithmic Framework for Deep Learning in Sports Video Classification.* With the basic framework in place, the first and foremost problem to be addressed during the research is the lack of experimental data. There are some publicly available sports video datasets, such as the Sport-1M dataset, which is currently the largest video classification benchmark dataset consisting of 1.1 million sports videos. Each video belongs to one of 487 sports categories, and this dataset does not tag the decomposed actions of a particular category of videos. In the current lack of professional sports datasets, to achieve an automatic description of free gymnastics videos, in this work, a free gymnastics decomposition action dataset containing videos of professional events such as the Olympic Games and National Games is constructed [21], and the description of videos is labeled according to professional commentary. For the discriminative power calculation of video frames, a number of methods have been proposed to solve this problem. One of the commonly used strategies is to use an attention mechanism. Therefore, in the approach of this paper, the attention mechanism is fused into the existing video description network to calculate the

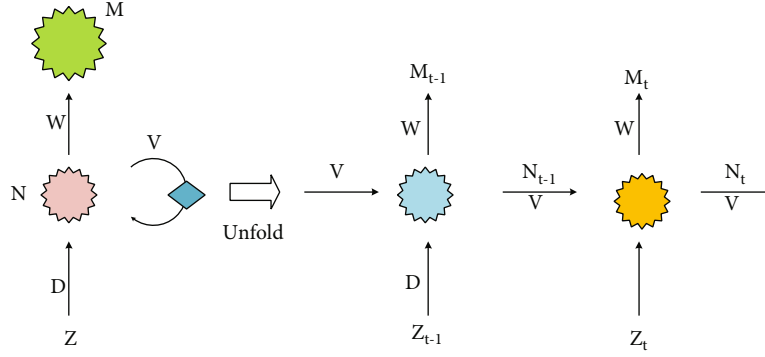


FIGURE 4: RNN hidden layer unfolding.

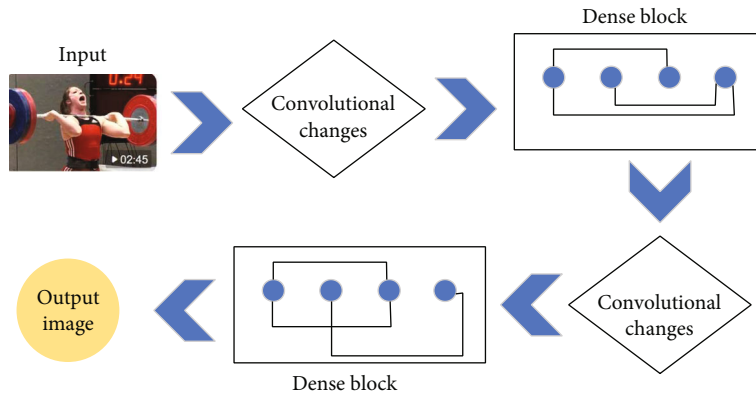


FIGURE 5: Basic flow chart.

weights between different video frames and improve the accuracy of the automatic video description. As shown in Figure 5, firstly, the free gymnastics decomposition action dataset is constructed; for the video data, feature extraction is performed by the convolutional neural network, and for the labeled text, a dictionary is constructed to extract the corresponding dictionary features; Figure 5 shows basic flow chart, then the training set is input to the video-to-text learning model for training, and the trained text that can accurately recognize the corresponding free gymnastics video is obtained; finally, the tested video features are input to the trained model to get the description of the free gymnastics video [22].

For the automatic description network of sports videos, the input data includes video sequences and text sequences. The video features are first extracted by a convolutional neural network, and then, the text features are extracted using natural language text processing.

**3.4.1. Video feature processing.** Convolutional neural networks have some degree of invariance to geometric transformations, deformations, and illumination. Trained convolutional neural networks can scan the entire image with a small computational cost and hence are widely used for image feature extraction. And while using its feature extraction, the specific morphology of the features is not considered at all.

**3.4.2. Description of text processing.** In this paper, we use one-hot vector coding to transform the descriptors of free gymnastics videos into features. The words in the annotated text of sports videos are first counted to construct a dictionary. One-hot computation process: the input is a sentence, and the output is a feature. The calculation is done by first calculating the total number  $N$  of all words described in the free gymnastics dataset and then representing each word as a  $1 * N$  long vector; in that long vector, there are only two values taken, 0 and 1. There is only one value, 1, in that vector, and the position where this value, 1, is located in the position of the word in the word list at the moment, and the rest of the values are 0. Its schematic diagram is shown in Figure 6.

## 4. Experimental Results and Analysis

**4.1. Experimental Setup.** The experiments in this section were done on the operating system application Ubuntu 16.04, and the code used to implement the experiments was based on the Tensorflow 1.6.0 framework, using the language Python 2.7. Its classification model diagram is shown in Figure 7.

The network model was trained on two NVIDIA Titan 1080 graphic cards with 11 GB of memory. The input video data is sampled at every 5 frames, the input to the C3D

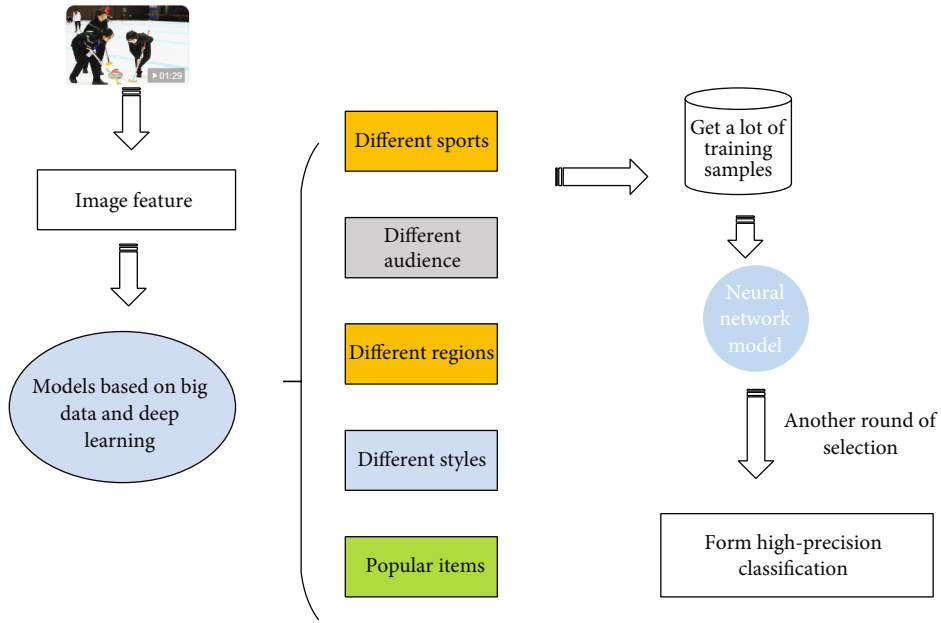


FIGURE 6: Codec structure diagram.

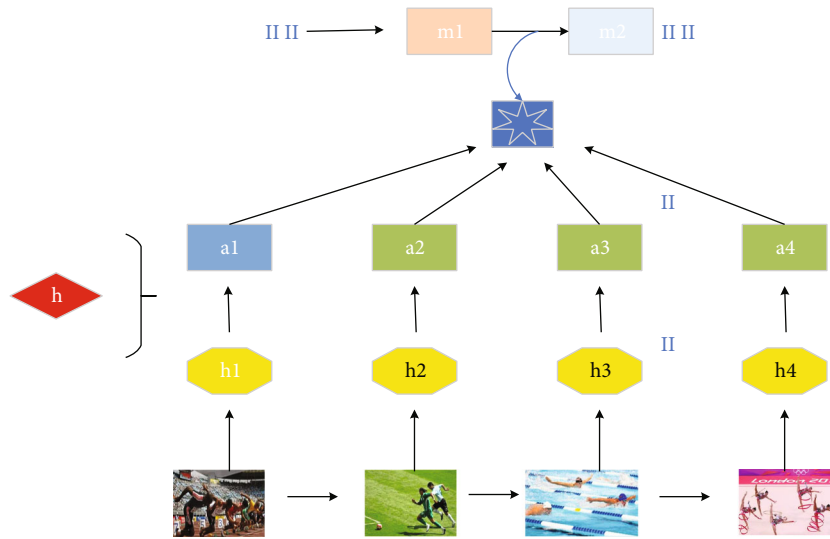


FIGURE 7: Sports-based video classification process.

feature extraction model is 16 frames long clips with 8 frames of overlap between two consecutive clips, and the fc6 activations of these clips are averaged to obtain 4096-dimensional video descriptors. This section uses LIBSVM classification tool, LIBSVM is a set of libraries for support vector machines, mainly used for classification, which needs to preprocess the fc6 learned video features in the format <label> <index 1>: <value 1> <index 2>: <value 2>.... (Table 2), for the databases that appear for each category.

This chapter is using a classification approach, so the data is described and categorized according to a sporting event professional actions; each video data consists of more than one decomposed action, so each video data is labeled as at least one category, and the categories are replaced with positive integers (starting from 1), for a total of 31 categories

[23]. Each category may require more than one word for its description, and the number of occurrences of this one category is shown in Table 2. It can be seen that half of the categories occur very infrequently below 10 times, with a relatively small number of categories occurring particularly high. Figure 8 shows an analysis of the frequency of the 16 categories that occur more than 10 times, with only a few categories having a particularly high number of occurrences.

#### 4.2. Analysis of Experimental Results

4.2.1. Analysis of Multicategorical Evaluation Indicators and Experimental Results. Accuracy is the most common performance metric in classification models and applies to binary classification models, but can also be used for



TABLE 2: Databases in which the video category appears.

Category	Quantity	Category	Quantity
1	23	1	34
2	33	2	35
3	21	3	32
4	34	4	12
5	32	5	43
6	34	6	47
7	21	7	55
8	21	8	42
9	32	9	33

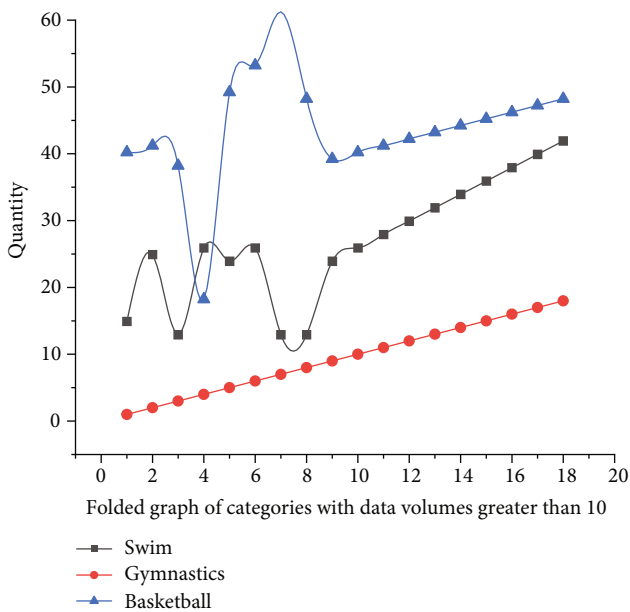


FIGURE 8: Folded graph of categories with data volumes greater than 10.

multiclassification models. The calculation of accuracy is also relatively simple. Assuming the classification model is  $g$ , its formula for calculating accuracy is as follows [24].

$$T(x) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{i=1}^n X_i^2. \quad (4)$$

4.2.2. *Example Analysis.* The experiments are representative frames of two videos from the test data of the self-built dataset, and due to the page limit, only two examples are given for reference; this example is the same video as the example in Section 3. The experimental results of the automatic description of free gymnastics on the self-built dataset with different models and the multilabel classification method in this chapter are compared. Compared to the original model, S2VT, the model with the attention mechanism is similar in the direction of the “forward” test in blue, but in the video, the improved model is more specific. In the classification problem, this video contains two categories, and the classifi-

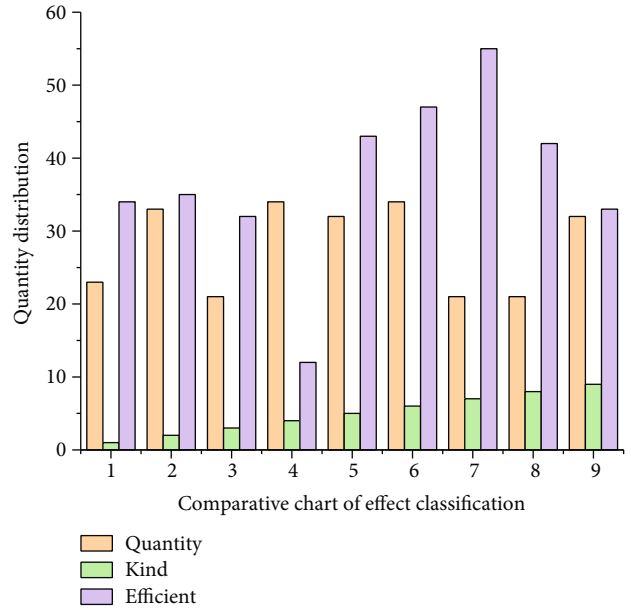


FIGURE 9: Comparative chart of effect classification.

cation results are significantly better. Figure 9 shows a comparison of the differences in the results after classification.

From the experimental results, it can be seen that the end-to-end deep learning algorithm proposed in this chapter for time-series accurate classification of sports video classification big data has a large improvement in the accuracy performance  $F$  compared to other algorithms for classifying sports video classification time-series data [25]. The LSTM-based model improves 14.22% over the BPNN model used on sports video classification and 3.91% over EE, the superior integrated learning algorithm in time series classification. The GRU-based model improves by 14.03% over the BPNN model used for sports video classification and by 3.72% over EE, the superior integrated learning algorithm for time series classification. However, in terms of classification time consumption, both LSTM and GRU have a large speedup over the EE algorithm, but both of them take 6.15 times and 6.03 times more time than the BPNN algorithm used for sports video classification, respectively [26]. The accuracy of these two algorithms can be applied to sports video classification large data temporal classification, but the operational efficiency can only be applied to nonreal-time sports video classification application scenarios.

## 5. Conclusion

In this paper, the methods and related technical theories of automatic video description and video classification are introduced, and the main framework and implementation steps of the free gymnastics video automatic description method based on long- and short-term memory networks and the sports video automatic classification based on support vector machine multilabel classification are described in detail, and the feasibility of the improved methods in this paper is verified by comparison experiments. In this paper, we first take the automatic video description method as an

entry point and review the current state of research on automatic video description methods, sports video research, and video classification. Then, from deep learning, we introduce its related concepts and development history, focus on describing the derivation and specific architectures of three important types of neural networks, and make structural dissection of typical network models, such as special recurrent neural network LSTM, respectively. Finally, the current state of research on feature extraction of video is analyzed, and the commonly used convolutional neural network feature extraction methods are classified and compared.

In this paper, an automatic description method of free gymnastics based on big data and deep learning classification is proposed. The research object sports video data is not only characterized by the presence of a large number of keyframes, but the definition of its high-level semantic things also is more fixed, and sports videos have certain choreography rules, so the automatic video description problem is transformed into a video classification problem, and the important technical support to promote this problem transformation is the existing video classification model with high accuracy. To preserve the temporal signal of the video, this paper uses a C3D feature extractor and feeds the extracted feature vectors into multiple binary SVM classifiers to accomplish the task of multilabel classification. Video classification based on big data and deep learning can bring users a better experience in the future and can better perform high-precision classification for sports videos. To verify the feasibility of the problem transformation, the classification results are mapped into natural language descriptions. The experimental results show that the classification model of sports video based on big data and deep learning can effectively improve the accuracy of sports video classification rapidly.

## Data Availability

The data used to support the findings of this study are included within the article.

## Conflicts of Interest

All the authors do not have any possible conflicts of interest.

## References

- [1] J. Wang, Y. Yang, T. Wang, R. S. Sherratt, and J. Zhang, "Big data service architecture: a survey," *Journal of Internet Technology*, vol. 21, no. 2, pp. 393–405, 2020.
- [2] C. Qi, "Big data management in the mining industry," *International Journal of Minerals, Metallurgy and Materials*, vol. 27, no. 2, pp. 131–139, 2020.
- [3] Z. Lv and L. Qiao, "Analysis of healthcare big data," *Future Generation Computer Systems*, vol. 109, pp. 103–110, 2020.
- [4] M. Gao, W. Cai, and R. Liu, "AGTH-Net: attention-based graph convolution-guided third-order hourglass network for sports video classification," *Journal of Healthcare Engineering*, vol. 2021, Article ID 8517161, 10 pages, 2021.
- [5] S. Shilo, H. Rossman, and E. Segal, "Axes of a revolution: challenges and promises of big data in healthcare," *Nature Medicine*, vol. 26, no. 1, pp. 29–38, 2020.
- [6] M. Zhao, A. Jha, Q. Liu et al., "Faster mean-shift: GPU-accelerated clustering for cosine embedding-based cell segmentation and tracking," *Medical Image Analysis*, vol. 71, article 102048, 2021.
- [7] W. Chu, P. S. Ho, and W. Li, "An adaptive machine learning method based on finite element analysis for ultra low-k chip package design," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 11, no. 9, pp. 1435–1441, 2021.
- [8] M. Zhao, Q. Liu, A. Jha et al., "VoxelEmbed: 3D instance segmentation and tracking with voxel embedding based deep learning," 2021, <https://arxiv.org/abs/2106.11480>.
- [9] M. A. Amanullah, R. A. A. Habeeb, F. H. Nasaruddin et al., "Deep learning and big data technologies for IoT security," *Computer Communications*, vol. 151, pp. 495–517, 2020.
- [10] R. H. Hamilton and W. A. Sodeman, "The questions we ask: opportunities and challenges for using big data analytics to strategically manage human capital resources," *Business Horizons*, vol. 63, no. 1, pp. 85–95, 2020.
- [11] S. Khanra, A. Dhir, and M. Mäntymäki, "Big data analytics and enterprises: a bibliometric synthesis of the literature," *Enterprise Information Systems*, vol. 14, no. 6, pp. 737–768, 2020.
- [12] H. Tamiminia, B. Salehi, M. Mahdianpari, L. Quackenbush, S. Adeli, and B. Brisco, "Google Earth Engine for geo-big data applications: a meta-analysis and systematic review," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 164, pp. 152–170, 2020.
- [13] M. Amani, A. Ghorbanian, S. A. Ahmadi et al., "Google earth engine cloud computing platform for remote sensing big data applications: a comprehensive review," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 5326–5350, 2020.
- [14] J. A. Leopold, B. A. Maron, and J. Loscalzo, "The application of big data to cardiovascular disease: paths to precision medicine," *The Journal of Clinical Investigation*, vol. 130, no. 1, pp. 29–38, 2020.
- [15] M. Holmlund, Y. van Vaerenbergh, R. Ciuchita et al., "Customer experience management in the age of big data analytics: a strategic framework," *Journal of Business Research*, vol. 116, pp. 356–365, 2020.
- [16] M. N. I. Sarker, B. Yang, Y. Lv, and M. M. Md Enamul, "Climate change adaptation and resilience through big data," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 3, pp. 533–539, 2020.
- [17] K. Yu, L. Tan, L. Lin, X. Cheng, Z. Yi, and T. Sato, "Deep-learning-empowered breast cancer auxiliary diagnosis for 5GB remote E-health," *IEEE Wireless Communications*, vol. 28, no. 3, pp. 54–61, 2021.
- [18] S. Min, B. Lee, and S. Yoon, "Deep learning in bioinformatics," *Briefings in Bioinformatics*, vol. 18, no. 5, pp. 851–869, 2017.
- [19] D. Ravi, C. Wong, F. Deligianni et al., "Deep learning for health informatics," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 1, pp. 4–21, 2017.
- [20] X. Hao, G. Zhang, and S. Ma, "Deep learning," *International Journal of Semantic Computing*, vol. 10, no. 3, pp. 417–439, 2016.

- [21] A. Taha, M. Alrabeiah, and A. Alkhateeb, "Enabling large intelligent surfaces with compressive sensing and deep learning," *IEEE Access*, vol. 9, pp. 44304–44321, 2021.
- [22] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals, "Understanding deep learning (still) requires rethinking generalization," *Communications of the ACM*, vol. 64, no. 3, pp. 107–115, 2021.
- [23] L. Lu, X. Meng, Z. Mao, and G. E. Karniadakis, "DeepXDE: a deep learning library for solving differential equations," *SIAM Review*, vol. 63, no. 1, pp. 208–228, 2021.
- [24] A. Echle, N. T. Rindtorff, T. J. Brinker, T. Luedde, A. T. Pearson, and J. N. Kather, "Deep learning in cancer pathology: a new generation of clinical biomarkers," *British Journal of Cancer*, vol. 124, no. 4, pp. 686–696, 2021.
- [25] C. P. Hensley, E. M. Lenihan, K. Pratt et al., "Patterns of video-based motion analysis use among sports physical therapists," *Physical Therapy in Sport*, vol. 50, pp. 159–165, 2021.
- [26] L. I. U. Ziyu, "Application of college basketball training teaching based on sports video analysis under network multimedia," *Solid State Technology*, vol. 64, no. 1, pp. 170–181, 2021.

## Research Article

# Innovation and Practice of Music Education Paths in Universities under the Popularity of 5G Network

Haoyu Cao 

Tangshan Normal University, Tangshan Hebei 063000, China

Correspondence should be addressed to Haoyu Cao; chy12348808@163.com

Received 15 August 2021; Revised 2 September 2021; Accepted 3 September 2021; Published 7 October 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Haoyu Cao. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the development of the 5G mobile Internet, cloud computing, Internet of Things, and other cutting-edge technologies, the era of big data has quietly arrived. The purpose of this paper is to explore the feasibility of the application of new technologies for music teaching in the era of big data in the context of the rapid development of science and technology in the information society, to enlighten and lead music teachers to apply the spontaneous and conscious awareness of new media and fully apply the new achievements of science and technology in the information society for future music classroom teaching, and to analyze the mode, method, trend, characteristics, advantages, and disadvantages of music teaching in the new media environment. The aim is to analyze the advantages and shortcomings of music teaching in colleges and universities and to find solutions and future development strategies for them so that in the future, the 5G Internet can better serve music lovers and better contribute to the cause of music education in colleges and universities. New media, as a product of constantly updated information technology, provides powerful data support for the development of various fields, and the education industry also needs new media to boost the rapid development of education information technology, which of course includes college music teaching. The effective integration of new media technology into the college music classroom can improve the classroom efficiency of music teaching with rich and diverse teaching resources and flexible teaching forms.

## 1. Introduction

With the development of modern science and technology, especially the rapid development of information technology represented by “5G Internet+,” modern teacher learning has been given a strong impetus. In the context of modern curriculum reform, teachers’ teaching ability has been put forward to new requirements. In traditional education, knowledge is mostly imparted by schools, and people have relatively few channels to acquire knowledge, but with the continuous development of “5G Internet+education,” there are more channels and ways to disseminate knowledge, and the way people receive education has also changed dramatically. The continuous development and deepening of “5G Internet+education” has gradually broken the monopoly of schools on knowledge dissemination, allowing education to move from closed to open. With the development of “5G Internet+education,” the place where people will receive edu-

cation in the future will not only be on the campus but more on the Internet platform, teachers will teach knowledge on the Internet, students will receive education on the Internet, and educational resources will be shared and spread on the Internet, which will realize the good development trend that real learning and online teaching complement and expand each other. The development trend is good. With the continuous development of “5G Internet+education,” so far, with the platform and advantageous resources of the Internet, music teaching has also emerged in many innovative modes such as catechism, mobile APP, and WeChat, which have laid a certain foundation for the development and research of music teaching. This has enabled the traditional classroom of music education to gradually break the limitation of time and space and realize the sharing of resources of excellent music teaching modes [1]. In such a context, the important issue facing school education is how to rationally analyze and actively respond to the new model of music teaching

under the background of “5G Internet+” and closely combine the traditional music teaching mode with the emerging music teaching to improve the quality of school education.

Under the influence of the Internet, the means of music teaching and delivery methods have produced significant changes, and the Internet will play a positive and beneficial complement to traditional music teaching, and effective compensations and useful new attempts can be made in the Internet-based teaching mode [2]. First, Internet teaching has a strong practical significance for making up for the temporality and spatiality of traditional music teaching. In the past, students who did not have the opportunity and conditions to learn face to face could have online music teaching in music teaching forums on the Internet. This kind of music teaching needs to have certain technical equipment, and through the instant transmission of video and audio, the limitations of space can be broken, and the Internet technology that we have today can fully support the implementation of this educational tool [3]. The second point is that we can collect the background and documentation of the works and the video and audio using the convenience of the Internet, which also contrasts with traditional music learning where finding information is very difficult and tedious. By searching on the Internet, we can access a large number of resources as well as materials without having to spend a lot of energy searching, sifting, and reviewing relevant materials, thus saving time and energy, allowing students to focus more on the learning process, to have more energy at their disposal for learning, and to have access to better quality video, audio, literature, and graphic resources.

## 2. Related Work

With the continuous improvement of network technology and the advent of the information age, music education has gradually begun to step into the era of networked innovation. As early as the 1990s, the United States began a bold attempt to informative education, applying IT technology to traditional teaching and learning development and penetrating various teaching fields. As a result, it also played a positive role in promoting educational reform in the United States after the new century. Under this influence, various countries in Europe followed suit and unveiled a new dimension of music education worldwide.

With the continuous development of primary and secondary education, teacher learning communities have received more and more attention from researchers, and under the influence of “5G Internet+,” researchers have associated them with teacher learning and explored new ways of teacher learning development. The literature [4] points out the constructive nature of “community,” which is a social organization and social connection with socially constructive characteristics, and its essential meaning is negotiation. Learning is a social culture of reflection in the meridian of the community, emphasizing the construction of the learner’s identity in the process of community activities. The literature [5] identifies ten characteristics of teacher learning as autonomous, experiential, purposeful, disciplinary, differentiated, staged, continuous, collaborative, contex-

tual, and planned. The literature [6] argues that teacher learning is as wide as a prairie and as numerous as trails, allowing for a variety of options. The literature [7] uses a virtual learning community, the online community teaching, and research, as a platform to systematically study learning communities, and makes a detailed analysis from knowledge connotation extension to learning models, as well as teaching applications. The literature [8] proposed a model for building virtual learning communities, taking the promotion of members’ knowledge construction as the starting point, and elaborated the process and characteristics of building virtual learning communities from macro and micro perspectives, including analysis from multiple intelligence perspectives, systematic construction of models, dynamic learning approaches, and learning theoretical foundations. In their study, the literature [9] analyzed in detail the value creation of building regional teachers’ online learning communities in the Internet era. In his study, literature [10] elaborated the theoretical basis and practical application of building teachers’ learning communities in the context of “5G Internet+,” and literature [11] proposed a specific O2O learning model for building learning communities in the context of “5G Internet+.” The literature [12] proposes a specific O2O learning model for building learning communities in the context of “5G Internet+,” which provides meaningful guidance for the construction of communities in practice. The paper [13] makes more people aware of the changing situation and educational changes brought by 5G Internet+ by grasping the time and situation of 5G Internet+education. The literature [14] analyzed the characteristics of music education in “5G Internet+” and judged the development trend of music education through research, which fully demonstrated that in the environment of “5G Internet+,” the geographical relationship of teachers’ learning community has been dissolved, and the membership relationship has become the most important one. These studies show that in the “5G Internet+” environment, the geographical relationship of teachers’ learning communities has been largely dissolved, and the membership relationship has become one of the distinctive features of teachers’ learning communities in the “5G Internet+” environment. Literature [15] points out that online music education will become the mainstream and dominant mode of music teaching in the future; literature [16] suggests that based on the rapid development of information technology and the growing public demand for music education, online piano education, which can provide more personalized teaching programs and educational services, will gradually become the mainstream trend in the development and advancement of piano education.

## 3. Innovation and Practice of Music Education Paths in Colleges and Universities under the Popularization of the 5G Network

*3.1. Innovative Model of Music Education in Universities under the Popularity of 5G Network.* Network information technology has distinctive features that are different from

traditional technology: first, fast information dissemination; second, open and free network; third, strong interactive two-way communication; fourth, powerful multimedia; fifth, huge information capacity; sixth, retrievable and easy to save; seventh, information has diversity; and eighth, information has uncertainty. In order to better apply network information technology, we must understand the characteristics of network information technology and the vein of integration with music teaching [17]. Under the premise of ensuring the understanding, the scope of application of network information technology in music teaching and the teaching efficiency can be maximized.

*3.1.1. Applied to the Teaching of Music Professional Theory Knowledge.* When music education was still in the traditional model, teachers mostly used the demonstration method or teaching method to explain music theory to students; i.e., the teacher sang or explained the music in the way to teach. This requires teachers to have good logical thinking and language skills as well as the ability to teach music and to strictly follow the standards of the teaching materials. Under the background of the modern network information society, the way and process of music teaching have changed, and the media tools can provide more pictures, audio, and video materials, making the teaching process more vivid and interesting, and the content is richer, which greatly enhances students' interest in learning. In addition, the amount of information in online media is extremely rich, and the information metabolism cycle is shorter so that people can grasp the latest theoretical knowledge of music in time to meet the needs of people's learning.

*3.1.2. Applied to Music Professional Skill Teaching.* At present, the content of music teaching in online media is very rich, and the teaching videos cover the actual classroom recordings of masters from various famous universities at home and abroad, i.e., the live recordings of music schools or famous musicians' performances. By watching these videos, music teachers can well discover the shining points of these masters with high attainments, to improve and perfect their teaching contents [18]. For students, these videos are invaluable textbooks that allow them to understand the charm of music and enhance their theoretical knowledge and practical skills through visualization. Because these videos are fully accessible, they can be downloaded or played anytime, anywhere, and can be paused or repeated according to their needs for better learning. The wide variety of music teaching software and music practice software available online allows teachers to improve their teaching efficiency, enhance their ability to express their views in the teaching process, and facilitate the implementation of music skills. For the emergence of new learning applications, music teachers can learn, become proficient in using, and participate in the design of targeted improvements to educational teaching software to get better teaching results.

*3.1.3. Online Vocal Teaching Breaks Through the Geographical Limitation of Classroom Teaching.* With the proficient of network technology, it is possible to make

online vocal teaching available anywhere, and students can arrange their learning progress anywhere and at any time according to their actual conditions and needs, solving the contradiction between learning and practice. By posting messages on the online platform to get help from professional teachers and other learners and listen to a wide range of opinions or suggestions, for example, by using the WeChat vocal teaching model, teachers and students can interact and learn and communicate in online groups. Students can not only get the maximum professional information and learn many aspects of professional vocal knowledge but also communicate through the public platform to solve the difficulties they encounter in vocal lessons.

As Figure 1 shows the constructivist theory support model, Kurt Lewin, a famous German psychologist, pioneered the term "group dynamics," which he defined as "a dynamic whole that shares the same developmental goals and is capable of interacting with each other." To study the interaction between group factors and the environment and the influence of group factors on individual factors, Llewyn proposed the famous field theory idea using the functional equation.

$$B = \bigcup_{i=1}^n X_i \gamma. \quad (1)$$

From this formula, individual behavior is influenced by both the internal conditions of the individual and the external environmental pressure in which the individual is placed. At the same time, changes in the behavior of individuals in this "group" may also lead to changes in the overall group dynamics. "The main manifestation of the positive influence of group dynamics on the individuals in the group is that the individuals form a relatively strong willingness to work together, reach mutual understanding and encouragement among them, and constantly carry out information sharing and knowledge exchange, which will significantly improve the work efficiency as shown in the following equation:

$$K = \bigcap_{i=1, j=1}^n \eta X_i Y_j + C. \quad (2)$$

The participants in the music teachers' learning community can form stable social relationships and create an atmosphere of common belonging through various interactive processes and then achieve mutual recognition and collaboration among the participants; the quality of community learning environment, in turn, influences the behavior of each participant, and in turn, the positive behavior of individual individuals can contribute to the gradual stabilization of the quality community atmosphere. In this way, the two affect each other and promote each other, thus realizing a benign structure of a common learning body. Therefore, group dynamics theory has the following important guidance for the development of music teachers' learning communities in the context of "5G Internet+." It is important to focus on the cohesion and centripetal force of the community: some scholars believe that social groups are not closed because they are static, but each individual is

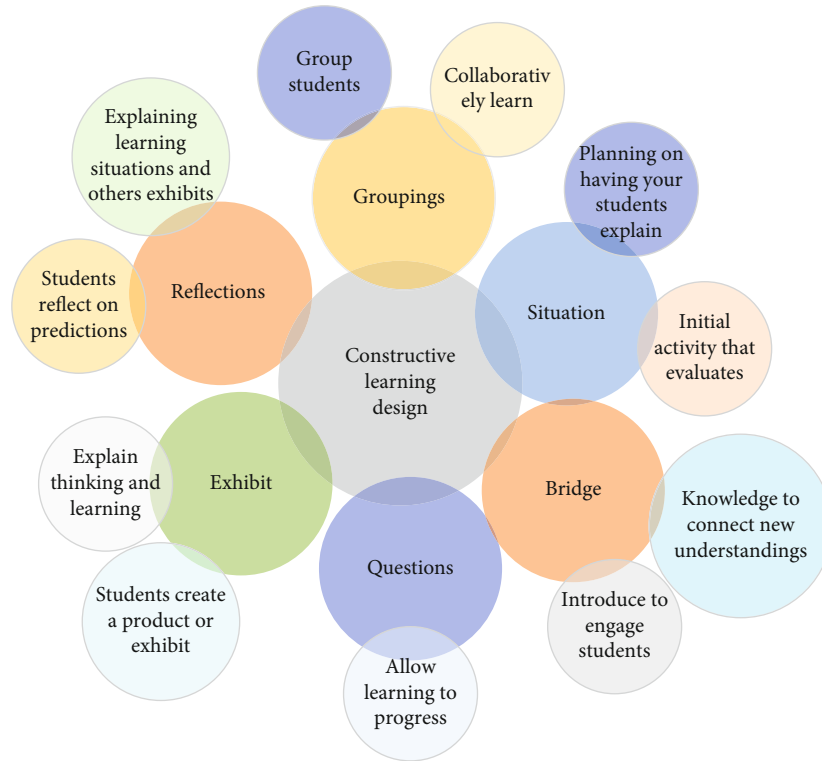


FIGURE 1: Constructivist theory support model.

interdependent, collaborative, and symbiotic because of their common learning or striving goals [19]. From the viewpoint of the group dynamics theory, the key point of music teachers' learning community formation lies in awakening the community consciousness of individuals, thus forming a common goal and vision, and strengthening members' sense of identity and belonging, to enhance the cohesion of the community and to promote excellence to form the community driving force: the incentive system can adopt quantitative settings, the exhibition of excellence platform, etc. to encourage members to actively participate; excellent members play an active role. The excellent members can play a leading role to drive the sustainable development of the community; the system should be used as a "whip" to play a binding role: the community under the vision of "5G Internet+" plays its characteristics of openness and interactivity, and at the same time, if the participants lack the awareness of relationship building or are too busy with their own learning needs, the community will not be able to develop. However, if participants are not aware of their relationship building or are too busy with their own learning needs and lack interaction and communication with other participants, they will get half the result with twice the effort. Therefore, the development of certain community norms, the implementation of relevant monitoring and management, and the formation of institutional constraints can enhance the sense of responsibility of each participant [20]. It is important to note here that the group dynamics theory effectively verifies the existence of significant influence of the environment on individuals; i.e., we must fully recognize the important influence of the external environment on individual

music teachers in the process of building music teachers' learning communities; update learners' relevant situations promptly with the help of members' learning situations, continuous statistics on relevant data, etc.; and carefully collect learners' suggestions, and specific interventions and adjustments should be given when appropriate. A model supporting the theory of group dynamics is shown in Figure 2.

Through cloud technology, massive amounts of knowledge, information, and technology can be saved, and students can communicate with teachers at any time, allowing teaching and learning to take place anytime and anywhere, removing the resource and space barriers between teaching and learning. Cloud technology not only solves the trouble of knowledge inventory in schools but also provides a quality teaching platform for education on the Internet. The development of an information-based teaching model requires the use of the Internet's technological routes, and "cloud technology" travels to meet this functional requirement, allowing for the rapid integration of campus network data routes to form a digital campus model. "Cloud technology" is a new storage model with similar functions to the Internet disk, but cloud technology is also a "new computing model" that can store amazing resources on its own, and it is a higher computing model formed on the Internet. The digital campus is the primary basic goal of "Internet+education," and now, some schools have completed the basic task of wireless network all over the campus and have started to promote the construction of "digital campus"; these facilities and exploration have laid the foundation for the development of "Internet+education." These facilities and explorations have

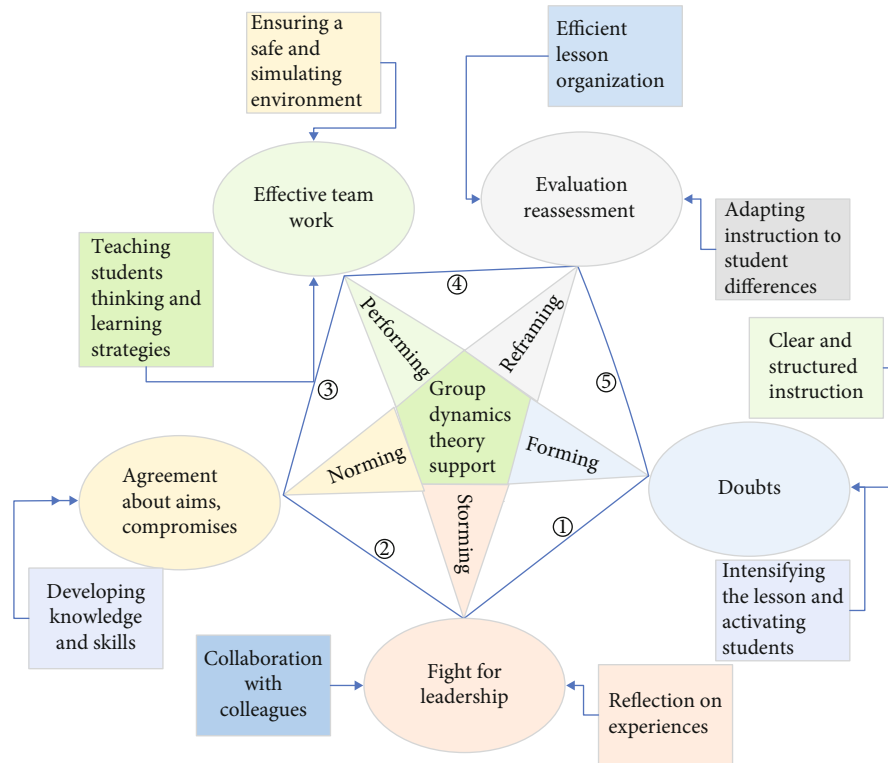


FIGURE 2: The group dynamics theory support model.

laid the cornerstone for the development of “Internet+education,” with “cloud+network+end” as the infrastructure of the Internet, teaching activities, students’ independent research and cooperative research, teaching resources, teaching materials, courseware, videos, and animation design as the means of teaching activities, the use of communication tools, face-to-face teaching video calls, and smart classrooms to set up the teaching environment, through big data survey of students’ age, gender, age, family background, and other elements. The teaching research based on “cloud technology” is shown in Figure 3. Comprehensive integration of educational resources, teaching research, and reasonable allocation are discussed.

The “Internet+music education” requires schools to change the perspective of education, with the help of cloud technology to make education transparent, visual, data-based, and intelligent; in this mode of communication and learning, it becomes the new orientation of teaching; cloud technology can quickly make accurate calculations of teaching resources that all rely on data to analyze the problem; this mode can make school management intelligent and convenient. With the emergence of talent, educational resource sweat, the development of cloud technology collection of these high-quality resources so that students can be dynamic on-demand, to ensure the centralization of teaching resources, intelligent. The campus is the carrier of education, and the development of education from the perspective of the Internet weakens the campus to strengthen the original purpose of education. This transformation of education helps education to a new level, which not only enhances the value of education but also allows a new level of integra-

tion of the Internet and education. “Internet+education” is not a single model, but a whole system model based on cloud technology that requires the collaboration of multiple systems to form an interconnected system, so Internet+education brings a team approach to teachers, teaching, and students. Music learning requires constant practice, constant eye opening, and constant acceptance of new teaching theories, but this increases the cost of learning. Knowing the world’s knowledge without leaving home has become the learning life that every scholar aspires to. Building an Internet platform has helped these scholars realize their dream of learning new theoretical information without leaving home or from anywhere. Through the Internet platform, students can learn at any time and teachers can teach at any time, realizing the requirement of truly free and idealized learning and breaking the limitation of time and geography.

3.2. *The Practice of Music Education Path in Universities under the Popularity of 5G Network.* Online education is a new segment of the education service industry that, despite its late start, is growing in size and influence at an alarming rate. The convergence of the Internet and the education industry has made it possible for us to receive knowledge in a more diverse and personalized way and without the constraints of time and space. Figure 4 shows that the Internet education market is expanding and growing at a fast pace and will continue to grow in the future. Along with the steady growth of the Internet education industry, the emergence of more new models and ideas will further diversify the industry. The rapid development of the online education industry has shown us the changes in the market and the



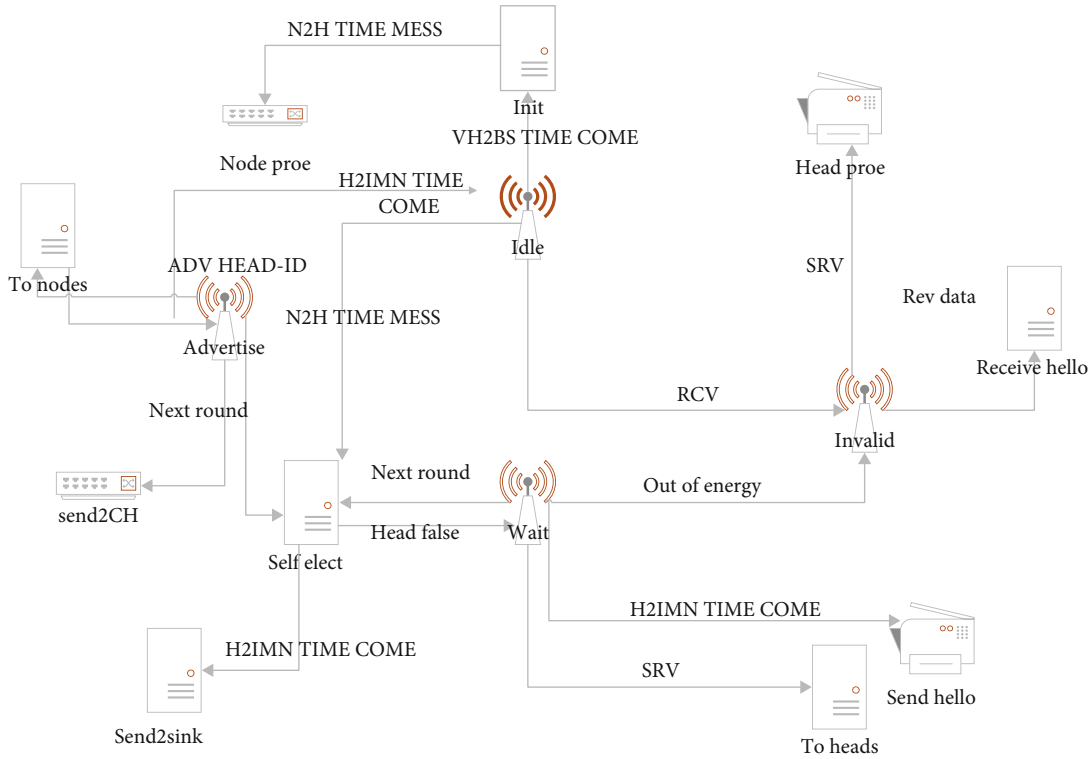


FIGURE 3: 5G Internet cloud technology-based teaching and learning research.

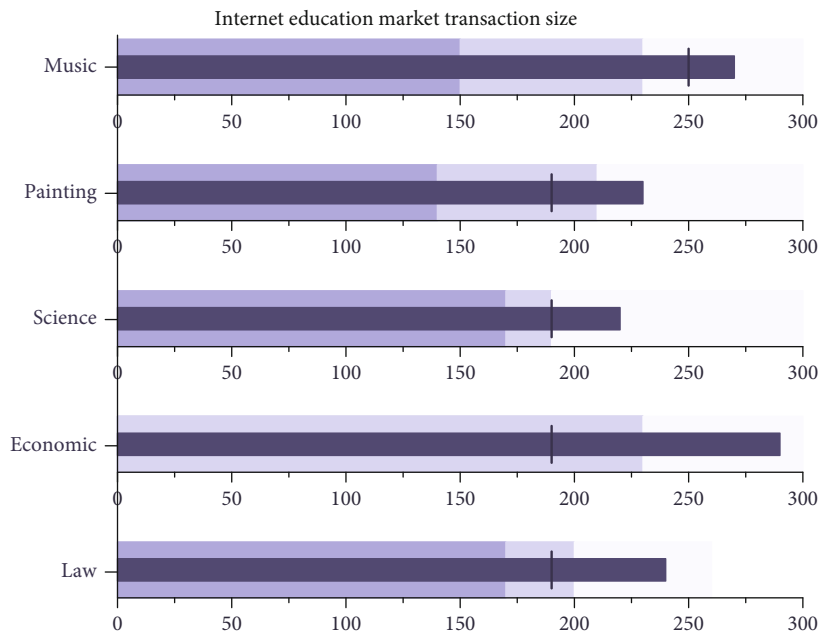


FIGURE 4: Internet education market transaction size.

direction of capital investment, and the development of the Internet education field is still a hot spot. From a macro point of view, there are deviations in education resources, and quality educational resources are mainly concentrated in a few regions, but given the advantages of Internet technology, education resources are used rationally and the problem of uneven distribution of education resources is

alleviated to some extent. At the same time, the demand of the mass market has led the online education industry to be subdivided into different fields based on the differentiation of content and audience groups. In the future, while exploring its operation mode, it is also necessary to make effective forecasts of future changes and control possible changes. The development of the Internet online education

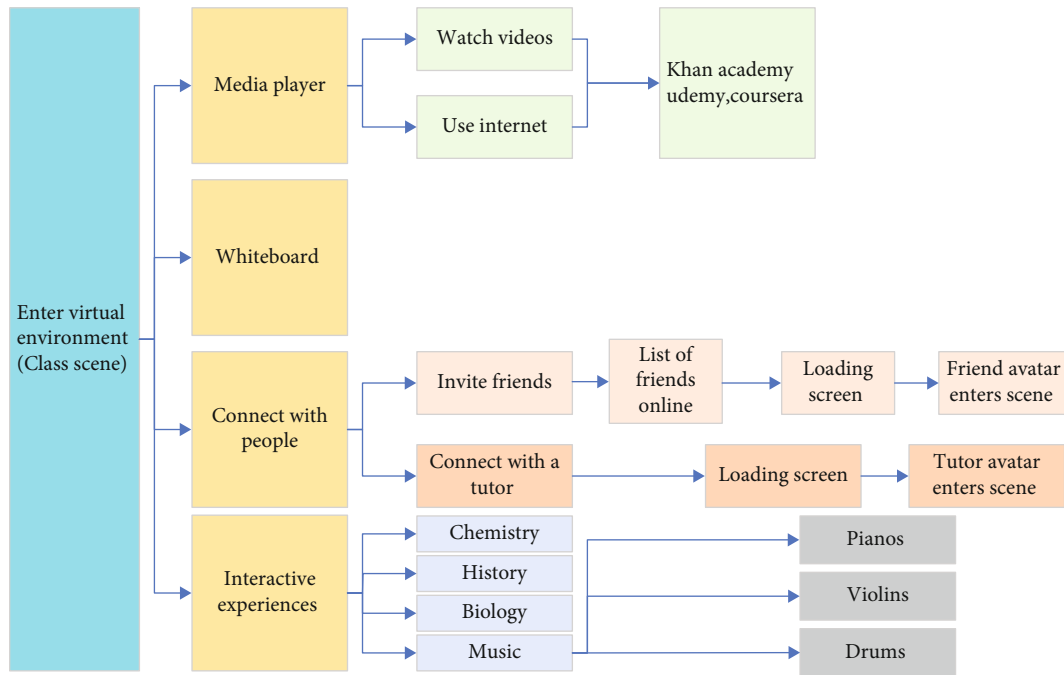


FIGURE 5: Virtual classroom uses case diagram.

industry is unstoppable, and the development of the arts in the education sector also has great potential for development, driven by the combined influence of national policies and economic and technological conditions.

In Internet education, teachers and students can be accommodated by creating virtual teachers to arrange education and teaching with teachers and administrators, while students study, communicate, take exams, etc. in the classroom. For virtual classrooms, first of all, students need to log in to the system, select the corresponding virtual classroom, and then apply to join it, and the teachers and administrators arrange the virtual classrooms. Each virtual classroom corresponds to one teacher. Virtual classrooms are opened by teachers, but in the backend of the system, administrators have the right to block some unusual classrooms. Once a virtual classroom is opened, the teacher user should fill in the relevant virtual classroom information. The teacher needs to set the information about the virtual classroom to be opened, such as the name of the course to be taught, the number of classes to be held, and the number of hours. After filling out the application, the administrator will review the application, and after passing it, the administrator will announce the class and the students will select the class. The virtual classroom is free to join and serves as a place for students to learn and communicate, but if you need to learn related courses, you need to provide points under the online education platform. According to the actual research, the demand for a virtual classroom of the online music education system is divided into classroom demand and student demand. The classroom requirement includes two functions: course content setting and student management, while the student requirement is to be able to freely choose the classroom. The specific use case description of the virtual classroom function is shown in Figure 5.

To promote the scientific development of Internet music education from the actual situation of Internet music education, the author designed and compiled a student questionnaire to understand and investigate students' musical quality background, school music activities, satisfaction with music courses, and other aspects of the questions. The research process chose the method of random distribution of questionnaires, and the questionnaires were selected for undergraduate students to ensure compliance with statistical theory and to minimize errors as much as possible. A total of 100 questionnaires were distributed, of which 100 valid volumes were returned. From the returned questionnaires, the students who participated in the survey came from different colleges and grades, and in a sense, the results of this questionnaire are still representative.

It is clear from Figure 6 that the observed (real known data) curve and the fitted (simulated data from the model) curve overlap, indicating a good model fit. The negative sentiment values are decreasing, while the positive and neutral sentiment values are increasing, indicating that people are not as skeptical and worried as they were when "Internet+music education" first emerged after one year of practice and exploration. Internet users began to think more about the problems encountered in the process of "Internet+music education" and showed a positive attitude toward "Internet+music education" and a more objective description of "Internet+music education." The description of "Internet+music education" also tends to be more objective. However, too many negative comments may be detrimental to the development of "Internet+music education," so the government needs to actively guide Internet users to think positively about the future of "Internet+music education" while worrying about the future of "Internet+music education," such as thinking about how to solve the problems

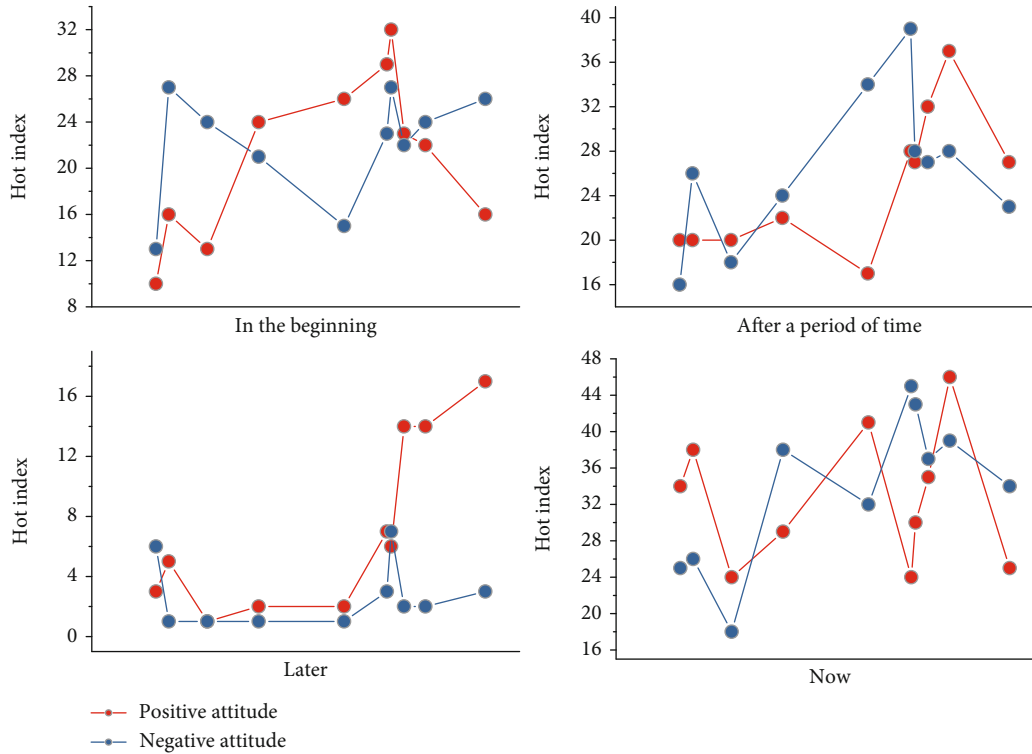


FIGURE 6: The changing hot index of Internet music education.

of “Internet+music education.” The government should guide the Internet users to think positively about the future of “Internet+music education,” such as how to solve the problems of “Internet+music education.”

Almost all of the students surveyed were interested in music. From Figure 7(a), 50% of the students like music very much, but among many music genres, they like pop songs as much as 73.34%, followed by classical music and instrumental music, accounting for 61.89% and 44.72%, respectively, Figure 7(a) shows that pop music is closer to the life of college students and is more popular among them, but there is no lack of students who like classical music and instrumental music and traditional music. Most of the students have a certain foundation of music knowledge and can read pentatonic and short scores, and they acquire music and related basic knowledge through different channels. 72.47% of the students learn music knowledge from music classes but also learn it by themselves, reading related books and new media on the Internet, accounting for 58.19%, 35.48%, and 47.94%, respectively; thus, it can be seen that students acquire music knowledge in more than one way. It can be seen that students have more than one way to acquire music knowledge. Since the Internet is still the second choice for students to acquire music knowledge and the types of music students like are more diverse, the content of music lessons should not be limited to present a more diverse music culture without losing national characteristics. As shown in Figure 7(b), 71.13% of students believe that music electives are necessary, while only 5.56% of students believe that they are not necessary and the rest are indifferent. This is mainly because most students believe that music education has the function of

cultivating emotion and regulating mood, 55.19%, and 39.26%, respectively, and that music education can also develop intelligence, enjoy entertainment, and improve the quality of life; 34.16% of students believe that music education has the above values. It can be seen that the students can clearly understand the functions and values that music has and agree with them, but they have some reservations about the music classes and music activities in school.

In addition, 74.26% of students feel that schools should offer online music electives because on the one hand, they want to improve their musical ability and appreciation through music electives, and on the other hand, they agree that music practice activities are useful for stimulating imagination, cultivating sentiment, and entertaining the mind. At the same time, 28.13% of students choose music electives out of interest in music, and 31.29% of students want to improve their overall quality through online music electives, although 26.35% of students still choose music electives because they think the exams are easy and they want to make up the credits. However, it still reflects that students’ awareness of the value of online music education is gradually increasing and they are beginning to attach importance to music courses. Nevertheless, 28.15% of students were dissatisfied with the current music elective courses, 52.94% felt average, and only 9.16% felt very satisfied. As seen in Figure 7(c), on the one hand, universities are gradually paying attention to online music education and continuously developing music courses and music activities, but due to the lack of publicity, the online music teaching process is not popular and students’ overall participation is not very motivated, resulting in most students not having participated in school music

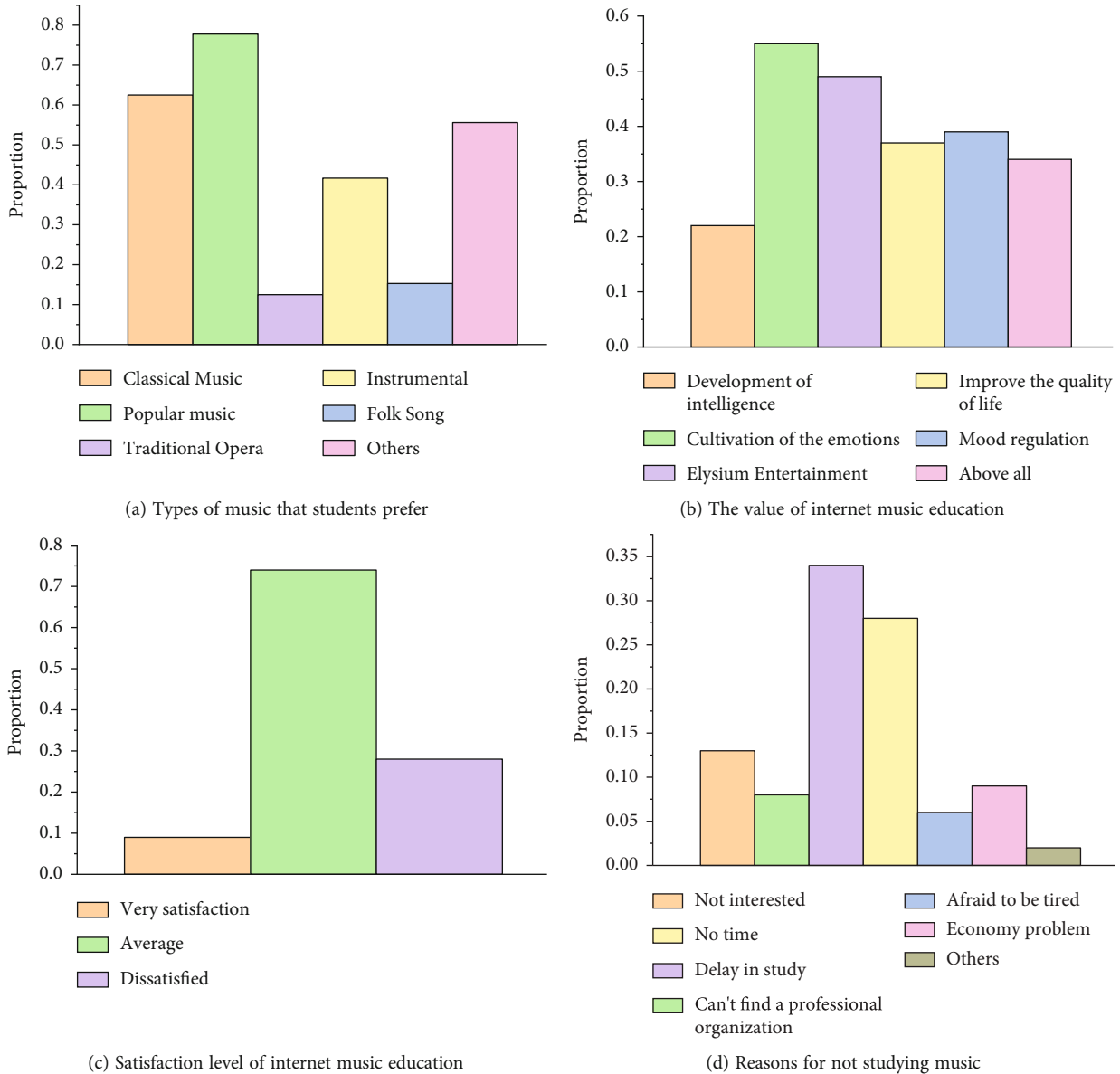


FIGURE 7: Results of student satisfaction survey on Internet music courses.

activities. In addition, most students who have not learned musical instruments, except for 13.75% of students because they are not interested in music, have not learned music because they are afraid of delaying their cultural studies (34.63%), they do not have time to learn (28.15%), and they cannot find professional institutions to learn music (8.24%), and difficulties in economic conditions (9.12%) also make some students not to learn music. Figure 7(d) illustrates that most of the students did not study music due to external conditions, and the university life is freer and has plenty of time; then, how to improve college students' interest in learning music is the focus of music education in comprehensive universities. Internet music education is a good solution to this series of problems, providing a new model as well as ideas for the path of music education in universities and providing more learning opportunities for students who like music.

Relying on Internet technology, students' developmental data such as physical and mental health information, social interaction information, and learning environment information at school can be integrated to enable their healthy physical and mental coursework development through music education. Internet education comprehensively accompanies the growth process and growth of students, and these data will accompany them through the student stage, and parents promptly discover the differences between students' different periods on the correct planning of their children's future development. The quality of teachers' services directly determines the improvement of teaching tasks in schools, and this series of reactions directly affects the development of music education. Schools to strengthen the construction of information technology and improve the development of education Internet all need the introduction of talent. The improvement of the overall teaching level of the school has

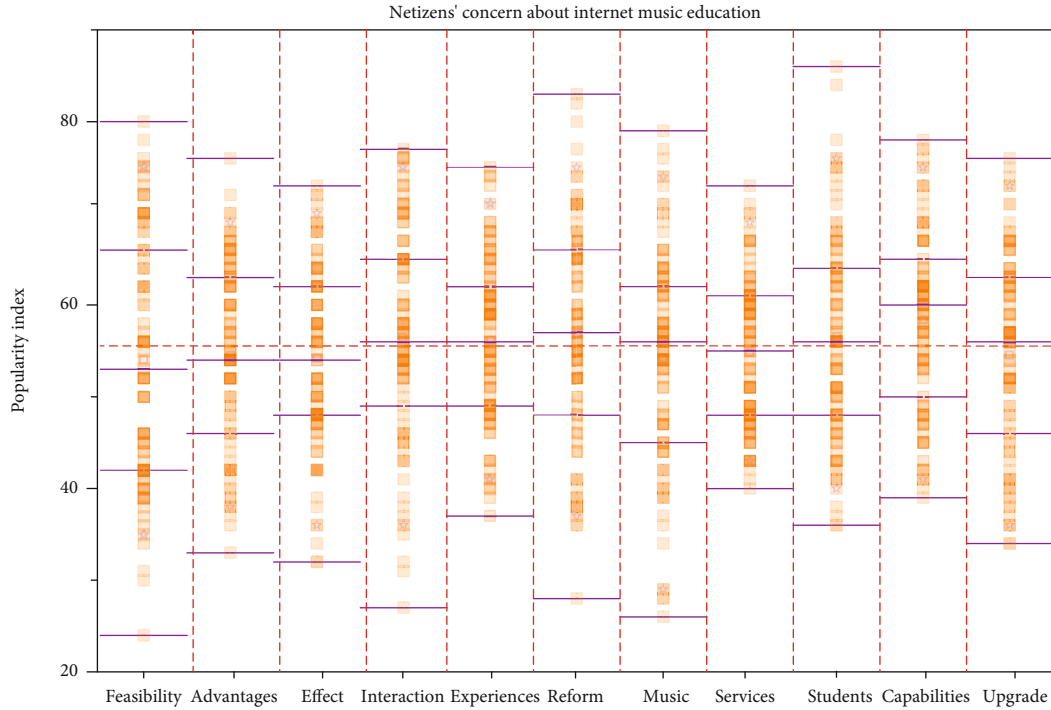


FIGURE 8: Internet opinion on music education on the Internet.

a great advantage for the recruitment of talents and the construction of the educational team. Information technology supports the holistic development of every student and ensures that every student can realize the value of life instead of surviving for the sake of survival. The Internet provides the basis for optimizing the educational and teaching environment, providing appropriate learning resources and effective learning support. The “Internet+music education” model has a huge mission to promote school construction and the holistic development of music teaching.

In this study, web crawlers and text tendency analysis techniques were also used to collect information on educational web opinion of “Internet+music education.” Web crawlers can crawl resources on the web based on keywords, depending on the efficiency of the program to determine the diversity of search content as well as the efficiency of updates, and is the essence of the search engine. When a web crawler opens a web page, the program collects information by analyzing the web page tags and obtains hyperlinks on the web page, and then, the search algorithm selects the next URL to be visited. The web crawler operates on the following principle: after starting the crawler program, it starts from the initial set of web addresses, resolves the DNS domain name, obtains the hyperlinks on the web page, and downloads this web page. The new hyperlink is added to the web address set, and then, the crawler crawls the next URL from it and so on until the crawler meets the end of the system. Text tendency analysis is to analyze and study the content of users’ opinions or comments on a certain matter or product to get whether the opinion or comment is a positive or negative attitude. Applying sentiment tendency analysis techniques in online opinion analysis can

uncover the attitudes and opinions of Internet users toward a certain social event. Text sentiment analysis can be used to understand the views, opinions, and attitudes of social people. The final results collected are shown in Figure 8, in which the words most closely associated with Internet music education are “feasibility,” “advantages,” and “influence,” and most people are concerned about the feasibility and advantages of Internet music education. This indicates that most people have a positive attitude toward Internet music education.

#### 4. Conclusion

The Internet is used in all aspects of our learning, education, and life. Many new models of vocal teaching have emerged through the Internet platform and information and communication technologies, and related teaching practices and academic research are continuing to heat up. With the advent of the Internet era, the discipline will encounter an unprecedented reshuffle, and the concept of correlation will be implemented in “big data+music education,” gradually connecting the barriers between the discipline of music education and computer, digital technology, and software programs, as well as in the teaching model with a new teaching and learning style that better reflects. “It will also provide students with personalized guidance and make it possible to teach according to their abilities and lifelong education. The combination of the Internet and music education is a demand for education, and the advantages of the Internet provide support for the integration of music education resources, which allows every student to enjoy the same educational resources, so it can be said that “Internet+music

education” supports the development of every student and promotes the overall development of every student; promotes the all-round development of every student’s knowledge, emotion, and mind; and ensures that every student can realize the value of life. The era of big data, as a specific stage in the development of the information society, has been surrounded by massive amounts of data, and its arrival has had a profound impact on education and even music education, and every educator has the responsibility to look at his or her profession with a futuristic vision. The integration of “Internet+” with the music education industry is also deepening further, as it gradually changes the way humans receive music education and drives the rapid development of the entire music education industry. Internet technology has driven a revolution in education, leading to an evolution of the college music education model, creating better educational opportunities for teachers and students as well as a learning environment that makes education more efficient and diverse.

### Data Availability

The data used to support the findings of this study are included within the article.

### Conflicts of Interest

The author does not have any possible conflicts of interest.

### References

- [1] O. Oyman, R. Koenen, P. Higgs, C. Johns, R. Mills, and M. O’Doherty, “Virtual reality industry forum’s view on state of the immersive media industry,” *SMPTE Motion Imaging Journal*, vol. 128, no. 8, pp. 91–96, 2019.
- [2] O. Golubchikov and M. Thornbush, “Artificial intelligence and robotics in smart city strategies and planned smart development,” *Smart Cities*, vol. 3, no. 4, pp. 1133–1144, 2020.
- [3] Y. Snihur, C. Zott, and R. Amit, “Managing the value appropriation dilemma in business model innovation,” *Strategy Science*, vol. 6, no. 1, pp. 22–38, 2021.
- [4] M. Muraszkiwicz, “The synergetic impact of AI, IoT, and 5G on information literacy and education,” *Zagadnienia Informatyki Naukowej—Studia Informacyjne*, vol. 57, pp. 7–22, 2020.
- [5] D. Cormier, P. Jandrić, M. Childs et al., “Ten years of the post-digital in the 52group: reflections and developments 2009–2019,” *Postdigital Science and Education*, vol. 1, no. 2, pp. 475–506, 2019.
- [6] D. Dash, R. Farooq, J. S. Panda, and K. V. Sandhyavani, “Internet of things (IoT): the new paradigm of HRM and skill development in the fourth industrial revolution (industry 4.0),” *IUP Journal of Information Technology*, vol. 15, no. 4, pp. 7–30, 2019.
- [7] D. Lewis, “Chapter chatter,” *IEEE Electromagnetic Compatibility Magazine*, vol. 8, no. 4, pp. 8–24, 2019.
- [8] S. Stickland, R. Athauda, and N. Scott, “Design and evaluation of a scalable real-time online digital audio workstation collaboration framework,” *Journal of the Audio Engineering Society*, vol. 69, no. 6, pp. 410–431, 2021.
- [9] J. Navarro-Ortiz, P. Romero-Diaz, S. Sendra, P. Ameigeiras, J. J. Ramos-Munoz, and J. M. Lopez-Soler, “A survey on 5G usage scenarios and traffic models,” *IEEE Communications Surveys & Tutorials*, vol. 22, no. 2, pp. 905–929, 2020.
- [10] A. Webb, “The 11 sources of disruption every company must monitor,” *MIT Sloan Management Review*, vol. 61, no. 3, pp. 65–70, 2020.
- [11] E. E. Bennett and R. R. McWhorter, “Digital technologies for teaching and learning,” *2020 Handbook on Adult and Continuing Education*, pp. 177–186, 2020.
- [12] A. Ghosh, “Commentary: new directions in the study of PRC-era science,” *East Asian Science, Technology and Society: An International Journal*, vol. 13, no. 3, pp. 443–448, 2019.
- [13] R. Frenneaux and A. Bennett, “A new paradigm of engagement for the socially distanced artist,” *Rock Music Studies*, vol. 8, no. 1, pp. 65–75, 2021.
- [14] R. Alt, A. Göldi, H. Österle, E. Portmann, and S. Spiekermann, “Life engineering,” *Business & Information Systems Engineering*, vol. 63, no. 2, pp. 191–205, 2021.
- [15] B. Wang, W. Mao, and G. Li, “China’s digital publishing moving towards in-depth integrated development,” *Publishing Research Quarterly*, vol. 35, no. 4, pp. 648–669, 2019.
- [16] M. Waqas, Y. Niu, Y. Li et al., “A comprehensive survey on mobility-aware D2D communications: principles, practice and challenges,” *IEEE Communications Surveys & Tutorials*, vol. 22, no. 3, pp. 1863–1886, 2020.
- [17] J. Mihailović, “The future of mobile operators-new business models,” *Management: Journal of Sustainable Business and Management Solutions in Emerging Economies*, vol. 24, no. 2, pp. 73–84, 2019.
- [18] P. James, R. Astoria, T. Castor et al., “Smart cities: fundamental concepts,” *Handbook of Smart Cities*, pp. 3–33, 2021.
- [19] J. Chen and H. Li, “Development prospect of China’s new consumer economy in the new situation—concurrently discussing the impact of COVID-19,” *Open Journal of Business and Management*, vol. 8, no. 3, pp. 1201–1205, 2020.
- [20] O. Oyman, R. Koenen, P. Higgs, C. Johns, R. Mills, and M. O’Doherty, “Virtual reality industry Forum’s view on state of the immersive media industry,” *SMPTE Motion Imaging Journal*, vol. 128, no. 8, pp. 91–96, 2019.
- [21] O. Golubchikov and M. Thornbush, “Artificial intelligence and robotics in smart city strategies and planned smart development,” *Smart Cities*, vol. 3, no. 4, pp. 1133–1144, 2020.

## Research Article

# Research and Implementation of the Sports Analysis System Based on 3D Image Technology

Hongwei Wang,<sup>1</sup> Jie Gao,<sup>2</sup> and Jingjing Liu <sup>3</sup>

<sup>1</sup>Capital University of Physical Education and Sports, Beijing 100191, China

<sup>2</sup>Beijing Sport University, Beijing 100084, China

<sup>3</sup>Capital Normal University High School, Beijing 100084, China

Correspondence should be addressed to Jingjing Liu; [liujj524@163.com](mailto:liujj524@163.com)

Received 23 August 2021; Revised 2 September 2021; Accepted 9 September 2021; Published 6 October 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Hongwei Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

On the basis of existing research, this paper analyzes the algorithms and technologies of 3D image-based sports models in depth and proposes a fusion depth map in view of some of the shortcomings of the current hot spot sports model methods based on 3D images. We use the 3D space to collect the depth image, remove the background from the depth map, recover the 3D motion model from it, and then build the 3D model database. In this paper, based on the characteristics of continuity in space and smoothness in time of a rigid body moving target, a reasonable rigid body target motion hypothesis is proposed, and a three-dimensional motion model of a rigid body target based on the center of rotation of the moving target and corresponding motion is designed to solve the equation with parameters. In the case of unknown motion law, shape, structure, and size of the moving target, this algorithm can achieve accurate measurement of the three-dimensional rigid body motion target's self-rotation center and related motion parameters. In the process of motion parameter calculation, the least square algorithm is used to process the feature point data, thereby reducing the influence of noise interference on the motion detection result and correctly completing the motion detection task. The paper gives the measurement uncertainty of the stereo vision motion measurement system through simulated and real experiments. We extract the human body motion trajectory according to the depth map and establish a motion trajectory database. For using the recognition algorithm of the sports model based on the 3D image, we input a set of depth map action sequences. After the above process, the 3D motion model is obtained and matched with the model in the 3D motion model database, and the sequence with the smallest distance is calculated. The corresponding motion trajectory is taken as the result of motion capture, and the efficiency of this system is verified through experiments.

## 1. Introduction

The technology that senses the motion of the human body through some sensors and can more accurately store and record it is computer sports capture [1]. The research fields of sports analysis technology based on 3D images include pattern recognition, computer image processing, computer vision, and computer graphics [2]. The application prospect of sports model technology based on three-dimensional images is very wide. It can be used as material in film, animation, game, and other systems. It can also save intangible cultural heritage in the form of sports for protection and has

real-time effects. The 3D image-based sports model technology can also be used in the real-time motion recognition field, such as somatosensory interaction [3]. Human body sports models based on three-dimensional images can be divided into two types: broad sense and narrow sense. The broad sense of motion capture generally includes the capture of facial expressions, gestures, and human bone joints, while the narrowly defined human body sports model based on three-dimensional images only refers to the capture of human bone joint motion [4]. With the gradual maturity of the development of sports model technology based on 3D images, there are already many methods to capture

motion, many of which have also been applied to actual projects (such as games, film and television animation, new generation of human-computer interaction, and action recognition) [5].

However, the currently widely used 3D image-based sports model system has many shortcomings. For example, the capture system has various hardware devices that need to be worn by athletes. These devices are expensive and complicated and require higher environmental requirements for the athletes. The accuracy is not high, and the user experience is not good. These shortcomings restrict the application of this technology in some fields to a certain extent [6]. In recent years, the depth camera has been developed rapidly, which provides a new method for the capture of motion. The 3D image-based sports model fused with the depth map is obtained because it avoids many shortcomings of the traditional capture system [7]. Under unstable lighting conditions, the sense of movement can be established by matching the obvious corresponding features in the space; that is, the visual moving target can be tracked in a long-distance space. Therefore, similarly, the method of tracking and analyzing specific markers in the image sequence can be used to detect motion [8]. The feature-based discrete measurement method is proposed based on this principle. It is suitable for measuring the motion parameters of long-term, large-volume moving targets, and the algorithm is relatively simple to implement. There are some effective linear algorithms, which require the measurement environment. It is also relatively low and is more suitable for field applications in industrial production and national life [9]. Because several traditional 3D image-based sports model technologies require the athletes to wear hardware devices or stick markers on their bodies and some also require athletes to wear special clothing, it is very inconvenient to recognize or interact with actions. In addition, these systems are often more expensive and require a special site, so it is difficult to achieve in general applications [10].

On the basis of the current mainstream 3D image acquisition method research, this paper proposes a method of using mode to acquire 3D images. At the same time, an iterative threshold method based on the depth value is proposed to remove the background of the depth image. Based on the characteristics of the existing 3D motion model reconstruction methods, a depth image-based 3D motion model reconstruction method is proposed. Through the use of the three-dimensional information in the depth image, the three-dimensional motion model can be easily restored. Based on the existing motion trajectory extraction technology, a method of first finding the joint points of the motion trajectory and then connecting the joint points with simple lines is proposed for a motion trajectory extraction algorithm. First, we calculate the distance between the corresponding components of the two sequences and traverse the generated  $m * n$  matrix from  $(1, 1)$  to  $(m, n)$  to get a path with the shortest distance. According to the process of this algorithm, we must first calculate the distance between the components in the sequence to measure the similarity of the sequence. For the measurement process and detection characteristics of the stereo vision three-dimensional motion measurement

system, the paper proposes an optimized design of feature marker rods as features. Each rod is composed of 5 features with strong reflective characteristics, which greatly improves the quality of feature imaging and the feature antinoise and background interference ability; it can simplify the algorithm of corresponding feature recognition and feature matching and ensure the accuracy of feature detection, feature extraction, and feature matching.

## 2. Related Work

The process of extracting the contour under the premise of the known image background will be simple and easy, but the environmental background is usually difficult to accurately determine, so there are many methods to propose a probabilistic model modeling method to estimate the probability of the background and the foreground to separate the contour [11]. In order to ensure the integrity of the contour, Li and Yang [12] used local features to encode the contour information and abandoned the use of the global features of the contour. A similarity measurement relationship can be established between them, so that the meaning of different human motion images can be recognized. In addition, Lu et al. [13] used device context to measure the similarity between images. The method has also better effect. When calculating the motion data, the two-dimensional information often cannot meet the requirements. Li et al. [14] considered restoring the two-dimensional information into three-dimensional information. For ordinary optical cameras, multiple cameras were required to shoot athletes from different angles to reconstruct the three-dimensional motion model of the human body. There are many methods based on this theory to conduct research, among which the method of three-dimensional object difference based on visual hull is different. The image sequence of the viewpoint was used to intercept the spatial cube, so that the real 3D motion model can be approached visually from the camera. Later, the method of Zhang [15] improved the three-dimensional object difference method, so that the camera cannot shoot the athletes at the same time. In addition, there is a method of reconstructing a three-dimensional motion model using depth values. This method is based on the principle of computer binocular vision. Two cameras are calibrated in advance to shoot the same athlete. The parallax of the two cameras is calculated to calculate the point on the image for depth value and then use the depth value to recover the three-dimensional motion model. In addition, some algorithms also take into account the timing relationship of the original image and combine the timing relationship for research. For example, Miyoshi et al. [16] measured the importance of each image edge in a moving image sequence which is the optical flow method.

Haralabidis et al. [17] described the structure of the human body with simple shapes and searched for the closest human body region in the input image to determine the body region. This process is the process of energy equation optimization. At present, there are many methods of using machine learning to recognize body parts. The recognition of various parts of the human body uses the machine



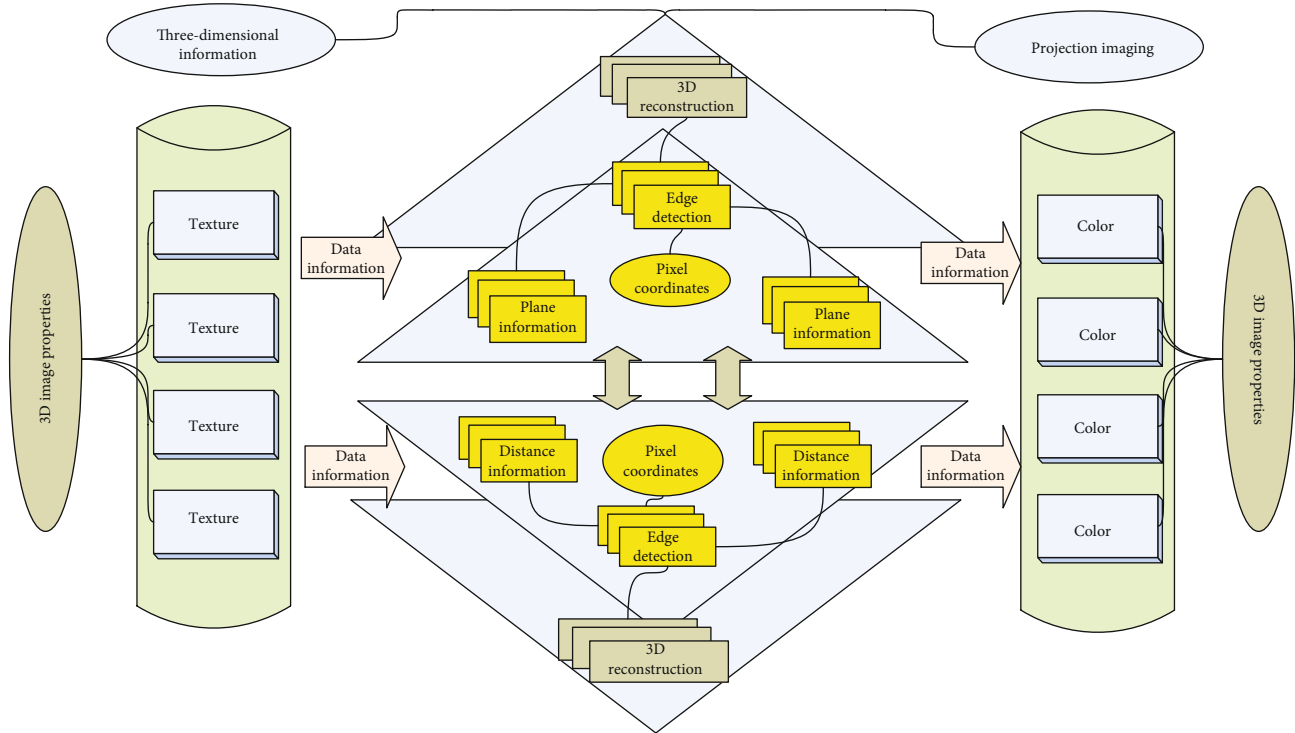


FIGURE 1: Hierarchical distribution of 3D image technology.

learning method based on AdaBoost. According to the image of human body movement, they find out the most easily recognized part according to the results of the previous human body part training and then find other parts step by step according to the inherent sequence of the connection of the human body. Modenese and Kohout [18] used a support vector machine to identify each part of the human target in the image and used a random decision forest method for marking body parts, which uses pixels instead of body parts for identification and analysis. The data set is trained to obtain a random decision tree, which is used to judge the attribution of pixels on the contour, and then, the pixels of each part are clustered to obtain the joint points of the human body. This method can quickly capture the movement of the human body, and Microsoft's latest somatosensory interactive device mode also uses this set of technical solutions. This type of discrete pose estimation method is based on body part recognition. Its advantage is that the motion recognition speed is faster and the algorithm efficiency is relatively high. However, because the joint points are estimated through the body part, there are errors and it is difficult to achieve the accuracy requirements for higher motion capture.

There are two main ways to optimize the model. One is optimization in a two-dimensional space. Some scholars use particle filters to optimize the created models. Particle filters have many and effective constraints on human motion. First, prescribing the athlete's initial posture (T-pose) is the action to be done at the beginning. The purpose of this is to make the local model and the athlete model easier to align and then use the current frame image to transform the local model and identify that particle filters are used to determine

other parts of the body [19]; the other way is to optimize in a three-dimensional space, and the local model established by scholars is represented by a three-dimensional point set. The depth data is transformed into a spatial point set. The matching is implemented between the spatial point sets, and the local model is further optimized. This matching process uses the nearest point matching method. The optimization of the local model is also iterative matching. The final result is to make the local model and the athlete's three-dimensional sports model similar [20]. The action captured by this fusion model method is smooth, and the quality is relatively good, but the algorithm has high complexity, and the real-time effect is difficult to achieve [21].

### 3. Construction of the Sports Analysis System Based on 3D Image Technology

**3.1. Hierarchical Distribution of 3D Image Technology.** Ordinary color images only contain data information such as the texture and color of objects, and the imaging principle of color images is projection imaging, so color images have no distance information of objects; that is to say, color images are two-dimensional, and it is difficult to form a three-dimensional space. The depth image has three-dimensional information. The pixel coordinates of the depth image represent the distance between the object and the camera (or depth sensor), which means that each pixel not only represents the plane information of each point  $(x, y)$  of the object. It shows the distance information perpendicular to the  $(x, y)$  plane, that is, the data information of the  $z$ -axis. Figure 1 shows the hierarchical distribution of 3D image technology.

Compared with the traditional process of extracting three-dimensional data from ordinary color images, the use of depth images is much simpler. The three-dimensional information of the depth maps can be used directly, which greatly simplifies the problems of three-dimensional reconstruction.

$$P(A, B) = P(A | B) \times P(B) = P(B | A) \times P(A). \quad (1)$$

Image data has a discontinuous characteristic, which is reflected by the edges of the image. For depth images, at the edges of the image, the depth values corresponding to pixels have changed. The edge information usually marks the end and the beginning of the area, and the area and the edge represent the basic image features, and many other features of the image can be obtained by deriving the basic features.

$$\begin{cases} A(i) \cap A(j) = \emptyset, \\ A = \cup A(i). \end{cases} \quad (2)$$

Sequence image motion detection technology based on the principle of stereo vision is a noncontact 3D measurement technology based on the principle of stereo vision and optical imaging, because of its noncontact, fast measurement speed, convenient and flexible measurement methods, and relatively high measurement accuracy. Higher advantages have been more and more widely used. The stereo vision motion detection system can generally be divided into the following modules: image acquisition module, feature extraction, recognition module, stereo matching module, camera parameter calibration module, spatial point positioning module, and motion parameter calculation module.

$$\text{cov}(X) = E\left[X \times X^T - \left(E(X) \times E(X)^T\right)\right]. \quad (3)$$

The early processing content in the field of image processing included edge detection. The characteristics of edge information include amplitude and direction. Along the trend of the edge curve, the pixels change smoothly, while when the trend of the vertical edge curve is vertical, the pixels change drastically, and this drastic change may be a slope shape or a step shape.

$$\begin{aligned} P(S(i), S(j), \dots, S(k) | T(t)) \\ = P(S(i) | T(t))P(S(j), \dots, S(k) | T(t)). \end{aligned} \quad (4)$$

In actual processing, the edge detection operator is often used to detect the presence or absence of an edge and its direction. The main operators for detecting the edge of the depth image are the Robert operator, Prewitt operator, and Sobel operator.

$$E(f(x)) = \frac{\sum_{i=1}^n \omega(t) \times f(x(t))}{\sum_{i=1}^n \omega(t)}. \quad (5)$$

Generally speaking, the computer vision motion detection method based on discrete features mainly includes three steps: the first step is to find the corresponding features in the image

sequence, and these features should have a certain uniqueness in order to distinguish and detect. The second step is to find the position of the corresponding feature in the next frame of a feature in the previous frame of the same image sequence, so as to complete the matching and correspondence.

$$E(f(x)) = \frac{\sum_{i=1}^n \omega(t) \times f(x(t))}{\sum_{i=1}^n \omega(t)}. \quad (6)$$

The third step is to design and give the motion model of the moving target, adopt an appropriate motion solution algorithm, and use the corresponding characteristic three-dimensional coordinates of each time point obtained by matching to calculate and solve the motion parameters and structural parameters of the moving target.

**3.2. Sports Model Architecture.** Figure 2 shows a schematic diagram of a sports model based on 3D images. A sports model system based on three-dimensional images usually includes three parts, namely, a transmitting source, a receiving sensor, and a data processing unit. The emission source is used to generate an electromagnetic field, and the distribution of the electromagnetic field is regular; the key parts of the athlete's body need to be equipped with receiving sensors (usually 10-20), and the connection between the sensors and the data processing unit is achieved through cables. In order to reduce interference and jitter, the sampling rate of this type of system is generally lower than 15 Hz, and for some high-speed sports such as football and basketball, such a sampling rate is far from meeting the requirements.

$$P(x) = \int P(B | A(x)) \times P(A(x)) dA(x), \quad (7)$$

$$\int f(x)p(x | t)dx - 1/n \times \int f(x)p(t | x)p(x)dx = 0.$$

The electromagnetic technology is mature, low cost, fast, and good real-time, and the calibration of the device is relatively simple. However, it has very high requirements on the environment. In order not to cause distortion of the electromagnetic field, metal objects cannot appear in the performance venue and nearby; otherwise, the accuracy will be affected.

The 3D motion model is first processed with other software tools before importing into Java3D, which is divided into the following three steps. The graphics simulation system requires a very precise shape, but it is not necessary for invisible internal features. The existing Pro/E model is used for design and processing, so it contains all the features of the part. Unnecessary internal features will cause the output file to be very large to directly output nearly three hundred files without making any changes. The files are about 17 MB and cannot be opened by the browser. Of course, they cannot be imported into Java3D. Therefore, the Pro/E model must be simplified. According to the virtual body, each object in the above picture can be extracted, and then, a meaningful name is given to each object, and the link is carried out according to the link relationship in the above

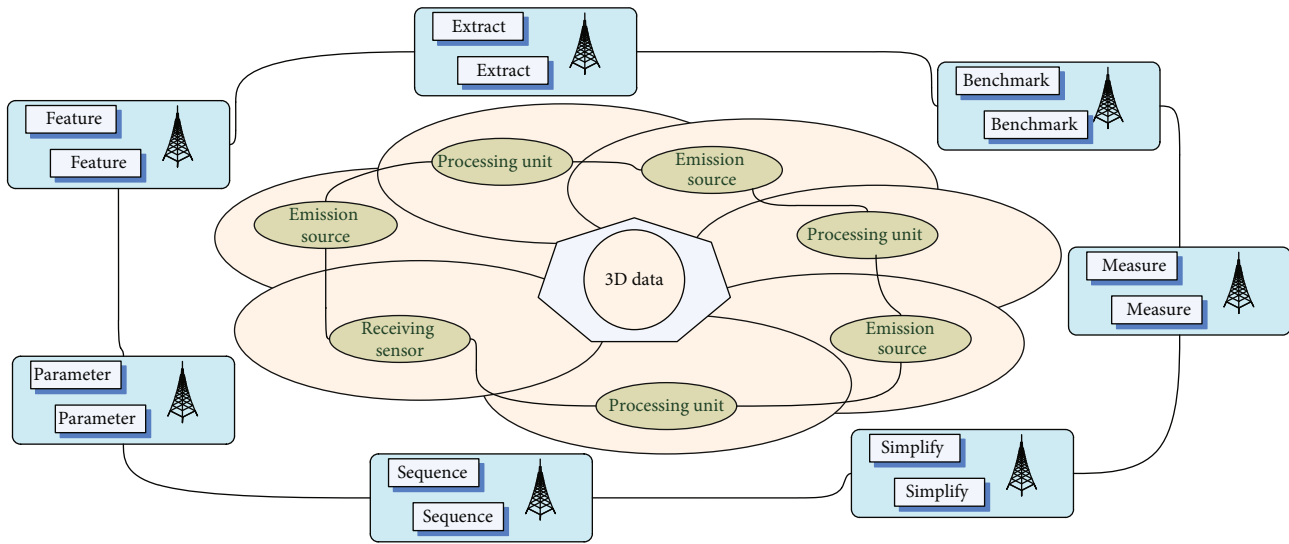


FIGURE 2: Schematic diagram of the sports model based on 3D images.

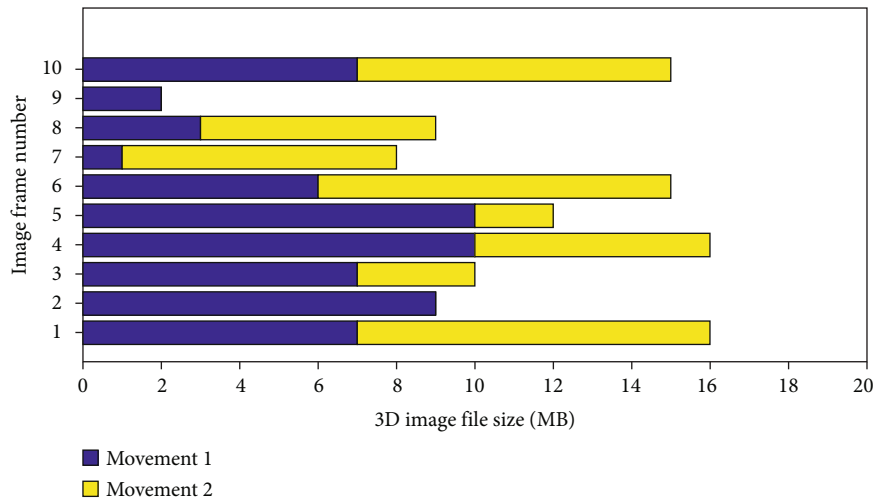


FIGURE 3: 3D image file size histogram of different motion processes.

picture. For the object that needs to be rotated, set the rotation axis, and you can easily move the rotation axis to a suitable position. There are mainly two ways to simplify: one is to remove the invisible parts and the other is to remove the invisible features of the visible parts. Invisible parts refer to the internal parts of the athlete and have no effect on the appearance. The invisible features of a part refer to the internal features of the part, such as some features such as holes inside the part. Figure 3 shows the three-dimensional image file size histogram of different motion processes. In the process of modification, we pay attention to the sequence of feature creation; otherwise, the error of not being able to find the benchmark will occur when opening it again. After simplifying the file size reduced to 4MB, the file has been reduced by three-quarters. Then, based on the simplified model file, it is applied to Java3D after some sorting.

The classic methods of three-dimensional distance measurement, such as Euclidean distance, can only be used to measure between a single frame, while a motion sequence

is composed of actions in multiple frames, and the length between sequences may be different, so the distance measurement method is not suitable for the measurement between sequences. By placing two or more predesigned feature marker rods composed of five feature marker balls with high reflective characteristics on the tested moving target, they will move with the moving target. The speed CCD camera is placed near the measured moving target, so that its field of view covers the entire range of movement of the feature marker rod and shoots the motion sequence image of the reflective feature marker ball to perform edge detection, center extraction, and center extraction on the feature target in the image. For feature matching and spatial point coordinate calculation, we design the target motion model, establish the motion parameter solving equation, and accurately solve its relevant 3D motion parameters and spin center. When the shooting rate of the camera is high enough, the trajectory of the spatial point can be obtained from the image sequence. Firstly, the system is calibrated, and then,

TABLE 1: Data cluster analysis of the sports trajectory model.

Data cluster number	Displacement (m)	Velocity (m/s)	Acceleration (m/s <sup>2</sup> )	Relative error (%)
1	0.98	1.12	-0.07	2.1
2	0.74	1.07	0.02	1.6
3	0.85	1.21	0.05	1.9
4	0.91	1.06	-0.03	1.7

the camera continuously photographs the athletes' movements and saves the image sequence. Then, the image sequence is processed and analyzed. The marker points are identified, and then, the spatial position of the marker points at each moment is calculated. If we want to get an accurate motion trajectory, the camera's shooting rate is required to be relatively high, at least 60 frames per second.

**3.3. Model Data Clustering Optimization.** Kinematics is a science that does not consider the forces and moments that produce motion in the research and specializes in the study of the laws of motion of objects. It involves the high-order derivatives of the position, velocity, acceleration, and position variables of a moving object with respect to time (or other variables). Athlete kinematics mainly analyzes the athlete's movement relative to a fixed reference frame as a function of time, especially the relationship between the joint variable space and the position and posture of the athlete's end effector. It includes the following two basic problems: (1) knowing the amount of motion of each joint (the angular displacement of the rotating joint, the linear displacement of the moving joint), where it is required to determine the position and posture of the end effector, which is the so-called positive problem of athlete kinematics (direct problem), and (2) determining the amount of motion of each joint according to the position and posture requirements of the end effector, that is, to solve the problem of how to make the end effector achieve the desired position and posture. This type of problem is called the kinematics of athletes' inverse problem (indirect location problem). Table 1 shows the cluster analysis of sports trajectory model data.

For trajectory planning in a three-dimensional space, it is necessary to specify the joint vectors of the athlete's starting point and ending point and then interpolate the joints to obtain the joint trajectory. The joint trajectory needs to meet a set of constraint conditions, such as the pose, velocity, and acceleration requirements of each node (start point, drop point, and end point), so that the joint position, speed, and acceleration are continuous in the entire time interval. VRML (Virtual Reality Modeling Language) files describe the abstract functional behavior of time-based interactive 3D multimedia information. The time-based 3D space described by the VRML file is called virtual world or realm for short. The graphic objects and auditory objects contained in it can be dynamically modified through a variety of mechanisms. The objects and their attributes in the realm are described by nodes, which form a scene graph according to certain rules. The first type of nodes in the scene graph is used to represent objects from the visual and auditory per-

spectives. The client program applies for a connection, and the server monitors all ports to determine whether there is a service request from the client program. When the client program requests to connect to a certain port, the server program connects the "socket" to the port. At this time, the server and the client establish a dedicated virtual connection. The client program can write the request like a socket, and the server program processes the request and sends the processing result back through the socket. They are organized in a hierarchical system and reflect the spatial structure of the realm. Figure 4 shows the hierarchical architecture of 3D motion image features. Another type of node participates in the event generation, and a routing mechanism is to form a route graph to determine how the realm changes dynamically over time.

Therefore, the feature-based motion measurement method can be summarized as extracting a set of discrete, sparse, highly discriminative two-dimensional features from the image, which correspond to the three-dimensional features of the moving target in the scene, such as points, lines, and surfaces. Then, based on the characteristics, the corresponding relationship between the sequence image frames at different moments is established. Under the necessary constraints such as rigidity assumptions, a set of targets containing targets are established according to the knowledge of motion dynamics, projective geometry theory, and the prior knowledge of related moving targets. The equations of motion parameters and structural parameters use the image coordinates (single-machine case) or space coordinates (dual-machine case) of the corresponding features between frames at different times to calculate and solve the equations containing the target motion parameters and structural parameters.

## 4. Application and Analysis of the Sports Analysis System Based on 3D Image Technology

**4.1. Feature Extraction of 3D Image Data.** We use the German SIMI Twinner software to find the synchronization point of the left and right cameras in the same process and then determine the starting point and ending point to be analyzed. The body and link center of gravity calculations are based on the Hanavan model which performs digital low-pass filtering and smoothing on the original data, and the filtering frequency is 6Hz. In the process of shooting, we shake the camera left and right to track the athletes. On the basis of ensuring that there are at least three control points in the background of each screen, try to make the images of the athletes large and clear, so as to facilitate subsequent video analysis and ensure the accuracy of analysis. In this experiment, assuming that the camera size is  $256 \times 256$  pixels, the illumination intensity function  $i(x, y)$  is based on the CCD camera imaging formula, and the camera coefficients are set to  $a = 0.5$  and  $b = 15$ ; the left half of the image is the reflected rate LR which is set to 0.1, and the reflectivity RR of the right half of the image is set to 0.15. The coordinate system of the image frame does not

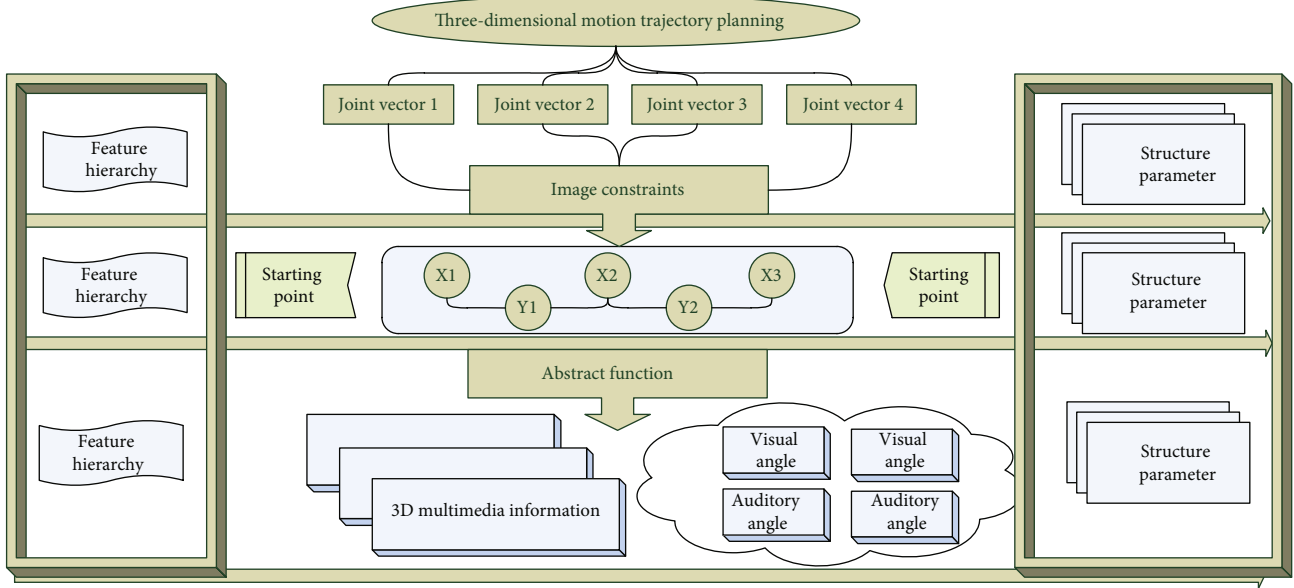


FIGURE 4: Hierarchical architecture of 3D motion image features.

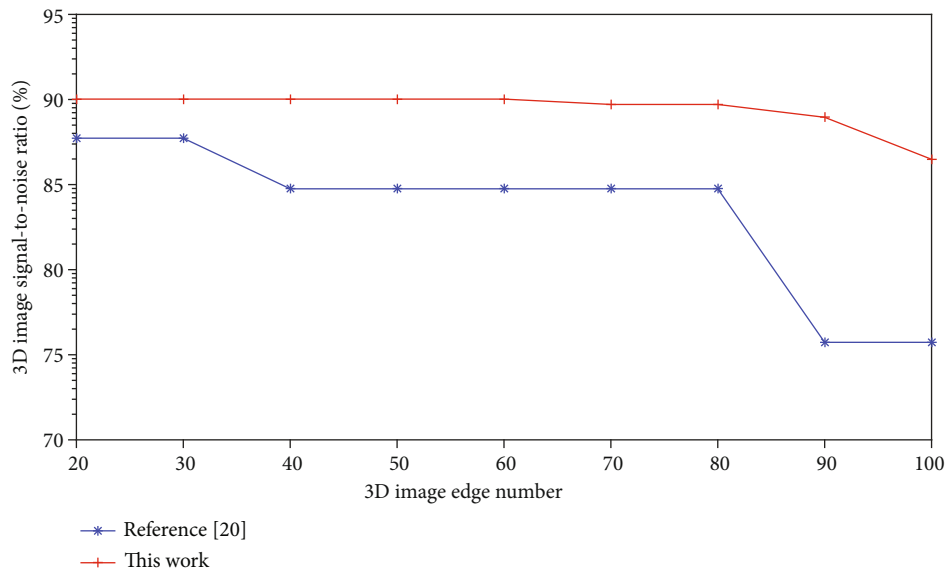


FIGURE 5: Comparison chart of the 3D image signal-to-noise ratio broken line.

represent the actual position coordinates of the visual field environment but represents the corresponding coordinates on the sensor, which are determined by the resolution of the sensor. From the above analysis, it can be seen that in order to restore the three-dimensional model from the depth image, the pixel coordinates of the depth image need to be converted into space coordinates, and the depth data is dimensioned, and the depth data needs to be normalized first. Subsequently, white Gaussian noise with  $t = 3$  is added to the image to synthesize a nonuniform weakly illuminated image for edge detection methods to detect the step edges in the image.

Figure 5 shows a comparison chart of the three-dimensional image signal-to-noise ratio broken line. The larger the signal-to-noise ratio is, the better the edge detec-

tion performance of the edge detection algorithm is. It can be seen that both methods have significantly larger signal-to-noise ratios because they both adopt the design idea of independent lighting. It can also be seen that in the low-grayscale and low-contrast noise-containing image area, the method fully considers the imaging model of the CCD camera and the illumination reflection image formation model, which overcomes the unevenness in the imaging process. As a result of the influence of lighting factors, it shows better edge detection and antinoise capabilities. We shoot the three-dimensional frame and determine the three-dimensional coordinates of the marker ball. The two cameras shoot the frame for about 2 seconds, pause, and then use the total station to measure the origin point  $a$  of the three-dimensional frame coordinates, point  $v$  in the  $x$ -axis

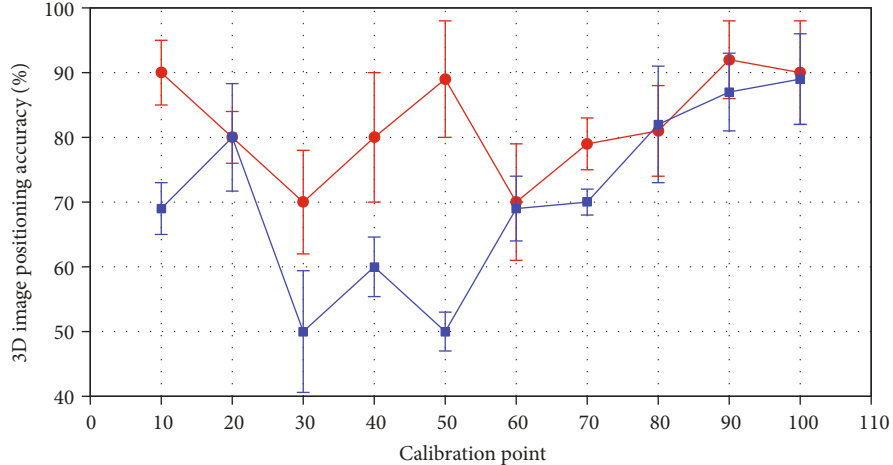


FIGURE 6: Statistical error distribution of 3D image positioning accuracy.

direction, and point  $j$  in the  $y$ -axis direction. For the coordinates of the additional control points, we use the software on the computer to collect the data observed by the total station and convert the coordinates of the additional control points into coordinates in the rectangular coordinate system defined by the three-dimensional calibration frame.

**4.2. Sports Sequence Simulation.** The main experiment in this paper uses two cameras to take pictures of the whole process of the three test jumps of the research object (the shooting frequency is 25 Hz), uses a total station to calibrate the peak frame and the additional control ball in a wide range of three-dimensional space, and uses the German belt Pan/Tilt/Zoom module function SIMI-Motion video analysis system for video analysis. In the analysis process, the human body digital model is used to perform digital low-pass filtering and smoothing on all the analyzed data, and the cutoff frequency is 6 Hz. The three-dimensional image analysis of human motion is realized by converting the two-dimensional image coordinates recorded on the film and video tape into the three-dimensional coordinates of the actual space. The direct linear transformation (DLT) algorithm directly establishes the relationship between the coordinates of the coordinate system and the coordinates of the object space. This linear transformation is achieved by taking a picture of the calibration frame and then calculating the photographic coefficient. The collected motion images are sorted, Zaziolski's human body model is selected, and the software is used for analysis. The data obtained by the analysis is smoothed by low-pass filtering, and the smoothed data obtained is analyzed, filtered, and sorted. The main points of this shooting method are to prepare a high-precision calibration frame. There must be at least 6 calibration points with known coordinates on the frame. The least squares method can be used to obtain the 10-element linear equations, which can be solved to obtain 10 photography and then find the three-dimensional coordinates of the point.

Figure 6 shows the statistical error distribution of the 3D image positioning accuracy. These data prove that the method has higher edge positioning accuracy than other

TABLE 2: Calibration coordinate system distribution of 3D moving images.

Image index	Track position	Track position	Track position
	X	Y	Z
1	0.12	0.04	0.04
2	0.07	-0.05	-0.08
3	-0.06	0.09	0.05
4	0.13	0.08	0.06
5	0.09	-0.07	-0.05

methods. The blue data line represents the control group, which is the accuracy of the model obtained according to the literature algorithm; the red data line represents the experimental group, which is the accuracy of the model obtained according to the algorithm proposed in this article. It can be seen from the comparison that the red data line is more stable and the displayed value is higher, which shows the superiority of the algorithm in this paper. This is because the improved wavelet multiscale multiplication edge detection technology and the fuzzy edge enhancement technology based on pixel gradient direction information are used in this paper, which greatly improves the antinoise ability and positioning accuracy of edge detection. In order to form the required trajectory in the three-dimensional space, the end pose is first converted to the joint vector angle through the inverse kinematics solution, and then, a smooth function is fitted to each joint, starting from the starting point and passing through all the path points in turn, finally reaching the target point. For each section of the path, each joint moves at the same time, which ensures that all joints reach the path point and end point at the same time. The three-dimensional space method describes the athlete's trajectory as a function of joint angles and performs trajectory planning. Table 2 shows the distribution of the calibration coordinate system of the 3D moving image. The three-dimensional space method does not need to describe the path shape between two path points in a rectangular coordinate system, and the calculation is simple and easy. And because the three-dimensional space and the rectangular coordinate space are not continuous

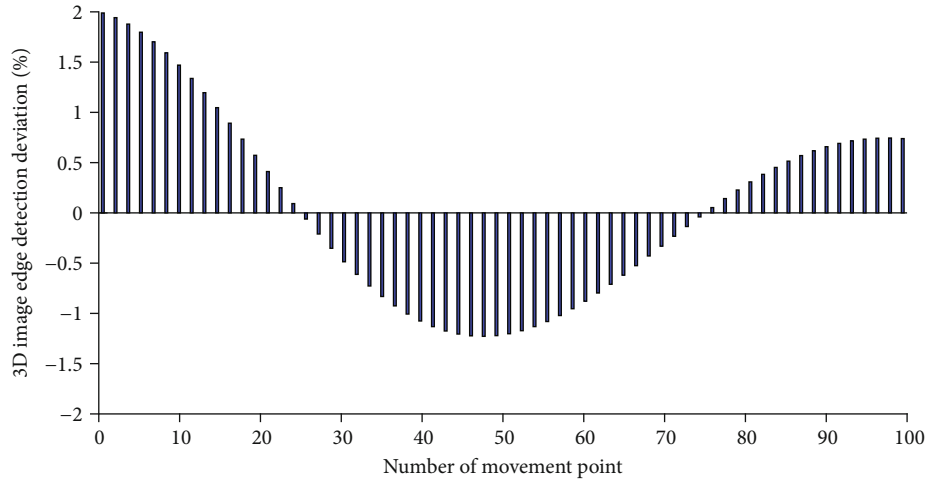


FIGURE 7: The distribution of deviations in 3D image edge detection of different moving points.

corresponding relations, the singularity problem of the mechanism will not occur.

We use this system to observe the control points and additional control points on each rod of the three-dimensional calibration frame, obtain their precise coordinates, and convert them to the designated coordinate system to track and scan three-dimensional images for a larger range of sports. The height of the center of gravity at the moment of take-off from the ground, the vertical distance from the moment of lift-off to the highest point of the sky, the distance from the highest point of the center of gravity to the crossbar after the take-off, and the center of gravity energy are given. The maximum height reached is not much different; the maximum height of the center of gravity for a successful test and two test jumps is not much different, but the vertical distance and the distance from the highest point of the center of gravity to the crossbar after the take-off are relatively large. Under the same coordinate reference system, the average difference of the absolute coordinates of the 15 points in the three-dimensional space measured by the total station and the image analysis is  $0.05 \pm 0.03$ , and the relative error of the three-dimensional tracking scan image measurement can reach 1.97%. Such measurement accuracy can meet the requirements of sports technical analysis.

**4.3. Example Application and Analysis.** We install the three-dimensional coordinate frame and place 100 additional marker balls along the athlete's approach route, each with a marker ball on the top left, bottom left, top right, and bottom right of the pole. According to the requirements of the SIMI Motion 3D scanning and tracking video analysis system, while the two cameras are scanning and tracking the moving target, it is necessary to ensure that there are more than two additional control points in the captured picture, so the additional control points placed in this experiment are confirmed by scanning and viewing by two cameras on the left and right, and there are at least two control points in the background of each screen. Although the method can accurately locate the edge pixels in some low-contrast areas, it cannot give better detection results for edges located

in low-contrast and low-grayscale areas at the same time. This is because the method does not consider the complete CCD camera imaging model and the illumination reflection image formation model, and its detection accuracy is affected by the absence in the CCD camera imaging formula, especially in low-grayscale areas. This influence greatly damages the quality of edge detection and positioning accuracy. Figure 7 shows the deviation distribution of 3D image edge detection for different moving points.

We see that in the case of nonuniform weak illumination, the edge detection accuracy of this paper is not greatly affected, and it can still give better edge detection results. The edges in the image can be positioned correctly, which further proves that the edge detection method is capable of detecting moving targets under nonuniform weak illumination, by creating the connected joints of the model, adding kinematics to the model, and assigning the three-dimensional coordinates of each joint of the human body analyzed in the kinematics method to the three-dimensional human body model to reconstruct the human body posture, perform balance analysis, and then analyze the established three-dimensional model in reverse dynamics. The joint will record the changes in the joint angle and muscle length when the model moves under the control of the motion guide point. From the data, it can be seen that the maximum height that athletes can reach is quite different, and there is a big difference in performance; there is no vertical distance between foreign athletes' body center of gravity from the moment of flying off the ground to the highest point of the sky. In terms of significant difference, the height of the center of gravity of the athletes before the flight is smaller than that of the world sports, and the difference is very significant; when the center of gravity reaches the highest after the take-off, the vertical distance between the center of gravity of the athletes and the horizontal bar is greater than that of the world athletes, and the difference is very significant.

We control the camera to perform 100 edge detections on the edge of the moving feature target at each movement speed at each exposure time, calculate the translation

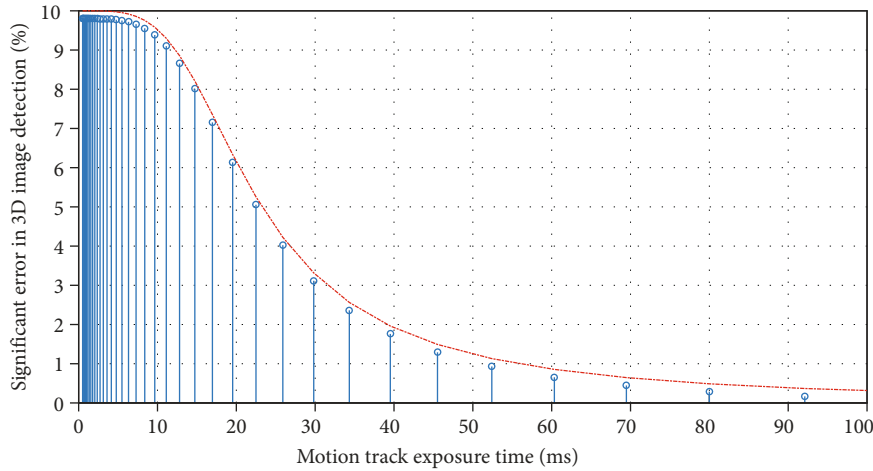


FIGURE 8: Three-dimensional image detection saliency error curve for different motion trajectory exposure times.

distance of the target, and give the mean value of the absolute error of the translation movement. Since the human body model and connection have been described in mathematical language and the computer system has been able to recognize it, after the partial modification of the parameters is input, the whole movement result will be changed to a certain extent, and then, it will be displayed in the computer for easy observation. It can be seen that when the exposure time is a constant of 40, 60, or 80 ms, initially when the moving target moves at a lower speed, as the target moving speed increases, the detection error does not decrease significantly. This is because at this time, the impact of the target's motion on the edge detection accuracy is not the main factor that affects the edge detection accuracy.

Figure 8 shows the three-dimensional image detection saliency error curve for different motion trajectory exposure times. Through many experimental verifications, in this motion detection test, when the motion speed of the moving target changes within 10-80 mm/s, the Dalsa CA-D6 camera produced by Dalsa in Canada and the reflective feature ball produced by Qualisys in the United States are used. When the target and the camera exposure time are controlled from 4 to 16 ms, the performance of edge detection is basically not affected by the change of target movement speed, and it has the best edge positioning accuracy. In general, in order to illustrate the reliability of the measurement, the reliability coefficient is usually calculated by mathematical statistics. The closer the coefficient is to 1, the more reliable it is, and the closer it is to 0, the less reliable it is. The reliability coefficient is slightly different due to different calculation methods, but it is generally believed that the reliability coefficient of group measurement should be above 0.70. At the same time, it is calculated that by increasing the number of light sources, lighting intensity, and reducing the exposure time of the CCD camera, the performance of edge detection can be unaffected by changes in the speed of the target in a larger range of target motion. With the further increase in the moving speed of the moving target, the edge positioning error caused by the target movement has gradually become the main influencing factor that affects the edge detection

accuracy and the translational movement distance detection accuracy. The accuracy of the edge detection gradually increases with the increase in the moving target speed.

## 5. Conclusion

Based on the 3D motion model of the depth image, this paper compares and analyzes the characteristics of the existing methods of 3D motion model reconstruction and then proposes a method of 3D motion model reconstruction based on the depth image. The depth image contains three-dimensional information, and the three-dimensional motion model can be easily restored by using the three-dimensional information in the depth image. The algorithm is used to realize the restoration of the 3D motion model, and the point cloud data of the 3D motion model is obtained. Through the introduction and analysis of the existing motion trajectory extraction technology, a motion trajectory extraction algorithm is proposed that first finds the joint points of the motion trajectory and then connects the joint points with simple lines. This algorithm avoids the problems caused by body self-occlusion and boundary noise and ensures the connectivity of the motion trajectory. Through the introduction of the existing 3D image-based sports model analysis methods, it is determined to use the recognition method to carry out the 3D image-based sports model simulation analysis of this article, and two one-to-one corresponding databases of the 3D movement model and the human body movement trajectory have been established. On the basis of the existing 3D vision measurement technology research, a stereo vision 3D rigid body movement and self-rotation center measurement method based on discrete feature marker rods is studied. For such a system, the focus is on stereo vision-based three-dimensional rigid body moving target motion modeling and motion parameter calculation, feature target edge detection and extraction in motion sequence images, corresponding feature matching between stereo sequence images, and circular feature target center extraction, and other key technical issues are considered. The motion parameters of the moving target's rotation,



translation, and spin center spatial positioning are measured, and the measurement uncertainty of the stereo vision motion measurement system is given, and the motion detection results of two groups of targets with different moving speeds are carried out. Through computer simulation, not only can the movement be analyzed according to the position, speed, angular velocity, and other kinematic parameters of the athlete when the movement is completed but also the movement can be adjusted according to the dynamic data such as the timing of the athlete's joint force, the amount of force, and the continuous working time of the muscles. Through analysis and comparison, simulation and real motion test experiments have proven the correctness of the motion model and corresponding motion algorithm proposed in this paper.

### Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

### Conflicts of Interest

All the authors do not have any possible conflicts of interest.

### Acknowledgments

Excellent talents training was supported by the Organization Department of Beijing Municipal Committee "Study on collaborative Development of Water Projects in Beijing-Tianjin-Hebei Region (No. 2016000020124G095)".

### References

- [1] X. Chen, "Framework design of sports image analysis system based on three-dimensional image technology," *Journal of Physics: Conference Series*, vol. 1982, no. 1, article 012208, 2021.
- [2] C. Ning, "Design and research of motion video image analysis system in sports training," *Multimedia Tools and Applications*, pp. 12–19, 2019.
- [3] Z. Y. Han, "The Wushu action design based on computer three-dimensional auxiliary system," *Journal of Discrete Mathematical Sciences and Cryptography*, vol. 21, no. 2, pp. 601–605, 2018.
- [4] C. Qu, "Virtual reconstruction of random moving image capturing points based on chaos embedded particle swarm optimization algorithm," *Microprocessors and Microsystems*, vol. 75, article 103069, 2020.
- [5] M. C. Tenovici, I. L. Petrovici, R. C. Vaduva et al., "Three-dimensional virtual model used to analyze a normal and prosthetic human joint," *Advanced Engineering Forum*, vol. 34, pp. 159–164, 2019.
- [6] S. Pu, "Development and application of sports video analysis platform in sports training," *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol. 23, no. 1, pp. 139–145, 2019.
- [7] X. Chen, "Development significance and structure analysis of sports video image analysis system based on VR," *Journal of Physics*, vol. 1982, no. 1, p. 2125, 2021.
- [8] C. Huang, Y. Zhang, C. Zhu et al., "Chinese sports basketball teaching tactics training system combined with multimedia interactive model and virtual reality technology," *Multimedia Tools and Applications*, vol. 2, pp. 13–15, 2019.
- [9] L. Xue, "Analysis of the sports teaching and training system based on the VR technology," *Frontier Computing*, vol. 18, pp. 1645–1650, 2020.
- [10] T. Ohmori, T. Kabata, Y. Kajino et al., "Three-dimensional limb lengthening after total knee arthroplasty in a simulation study," *Modern Rheumatology*, vol. 28, no. 6, pp. 1029–1034, 2018.
- [11] H. Wang, "Three-dimensional image recognition of Athletes' wrong motions based on edge detection," *Journal Européen des Systèmes Automatisés*, vol. 53, no. 5, pp. 733–738, 2020.
- [12] F. Li and X. Yang, "Correction simulation of three dimensional image of table tennis players serve position," *Information Processing and Advanced Education*, vol. 4, pp. 6–10, 2020.
- [13] N. Lu, Y. Wu, L. Feng, and J. Song, "Deep learning for fall detection: three-dimensional CNN combined with LSTM on video kinematic data," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 1, pp. 314–323, 2019.
- [14] J. Li, M. A. Marra, N. Verdonshot, and Y. Lu, "A three-dimensional finite-element model of gluteus medius muscle incorporating inverse-dynamics-based optimization for simulation of non-uniform muscle contraction," *Medical Engineering & Physics*, vol. 87, pp. 38–44, 2021.
- [15] L. Zhang, "Evaluation and simulation of sports balance training and testing equipment based on medical video image analysis," *IEEE Sensors Journal*, vol. 20, no. 20, pp. 12005–12012, 2020.
- [16] E. Miyoshi, T. Takaki, M. Ohno, Y. Shibuta, S. Sakane, and T. Aoki, "Large-scale phase-field simulation of three-dimensional isotropic grain growth in polycrystalline thin films," *Modelling and Simulation in Materials Science and Engineering*, vol. 27, no. 5, p. 054003, 2019.
- [17] N. Haralabidis, G. Serrancoli, S. Colyer, I. Bezodis, A. Salo, and D. Cazzola, "Three-dimensional data-tracking simulations of sprinting using a direct collocation optimal control approach," *PeerJ*, vol. 9, article e10975, 2021.
- [18] L. Modenese and J. Kohout, "Automated generation of three-dimensional complex muscle geometries for use in personalised musculoskeletal models," *Annals of Biomedical Engineering*, vol. 48, no. 6, pp. 1793–1804, 2020.
- [19] H. Ba, "Medical sports rehabilitation deep learning system of sports injury based on MRI image analysis," *Journal of Medical Imaging and Health Informatics*, vol. 10, no. 5, pp. 1091–1097, 2020.
- [20] J. P. M. Tribst, A. M. O. Dal Piva, A. L. S. Borges, and M. A. Bottino, "Simulation of mouthguard use in preventing dental injuries caused by different impacts in sports activities," *Sport Sciences for Health*, vol. 15, no. 1, pp. 85–90, 2019.
- [21] G. J. Tierney, H. Joodaki, T. Krosshaug, J. L. Forman, J. R. Crandall, and C. K. Simms, "Assessment of model-based image-matching for future reconstruction of unhelmeted sport head impact kinematics," *Sports Biomechanics*, vol. 17, no. 1, pp. 33–47, 2018.

## Research Article

# Research and Development of Inventory Management and Human Resource Management in ERP

Bo Zhao <sup>1</sup> and Chunlei Tu<sup>2</sup>

<sup>1</sup>Chengdu Sport University, Chengdu 610041, China

<sup>2</sup>Chengdu Textile College, Chengdu 611731, China

Correspondence should be addressed to Bo Zhao; [cnszhaobo@163.com](mailto:cnszhaobo@163.com)

Received 24 August 2021; Revised 16 September 2021; Accepted 17 September 2021; Published 5 October 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Bo Zhao and Chunlei Tu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Nowadays, more and more companies are applying total inventory management as well as human resource management as one of the core concepts of the enterprise management platform. ERP is a resource allocation platform based on information technology applications to have the advantage of advanced and comprehensive management ideas to provide planning and operation software for enterprise managers and employees. This paper describes in detail the ERP system, as well as its planning and control ideas, its ideas, and the idea of internal control of enterprises in line with the rapid development of the information economy and knowledge economy in today's situation; ERP systems also have some inefficiencies and other situations; in response to this situation, this paper also analyzes the ERP system by analyzing the application of ERP systems in enterprises, from the market function information system and ERP system in the application to inventory management as well as to human resource management for decision-making, exploring its deficiencies in the inventory management system and human resource system and then proposing corresponding improvement methods and corresponding development measures. Enterprises can use the improved advanced information technology to enhance the unity and sharing of data and fine management and improve operational efficiency, and enterprises can achieve standardized and process-oriented management of daily operations by using the improved ERP to control various business processes such as production, procurement, and sales and realize the collaborative processing of financial and business processes.

## 1. Introduction

Global economic integration is a major trend in international economic development which in recent years has brought rare opportunities and serious challenges to enterprises. To win in the competition, enterprises have to improve their shortcomings in a lot of ways. The price of product enterprises was reduced on the one hand through a variety of effective management tools and technology to cut costs and on the other hand through the aspect of using effective enterprise information management tools [1]. In China, with the development of the market economy are modern enterprises. The ERP system is the most advanced enterprise management technology and method that emerged in foreign countries at the beginning of the century

and can bring great benefits to enterprises. In this paper, inventory management in the system is the object of study. In the framework of the supply chain, inventory management is the most critical link in the special management of manufacturing enterprises; it is no longer limited to simple material resource allocation and management, which puts forward higher requirements for inventory management. With the advent of the information age, the ERP system has become the mainstream means to improve the management level of a lot of enterprises and has been widely used in supply chain inventory management [2]. In the ERP environment, modern network technology and computer platform are used to build an information system to achieve efficient management of supply chain inventory, although it can collaborate with all aspects of the procurement,

production, and sales system; at the same time, it also faces new problems, highlighted in inventory control, which affects the effectiveness of the implementation of the whole ERP system, so the study of supply chain inventory management in the ERP environment has become a new subject. However, with the continuous development of economic globalization and the increasingly fierce market competition, especially the impact of the financial crisis, the OEM enterprises at the end of the industrial value chain are facing a serious situation due to the backlog of inventory and poor circulation of production materials. In the marketing category, inventory can be said to be an important part of enterprise logistics; the potential for inventory cost reduction is much greater than other marketing links, because logistics costs cover almost half of the enterprise marketing costs and about 30% of the total value of the product, while inventory costs account for 35%.

The ERP system is the abbreviation of Enterprise Resource Planning. It is based on information technology and integrates information technology and advanced management ideas. With systematic management ideas, it provides decision-making methods for enterprise employees and decision-makers. For the management platform, it is a new generation of integrated management information systems developed from MRP (Material Requirement Planning). It extends the functions of MRP, and its core idea is supply chain management. It goes beyond the boundaries of traditional enterprises, optimizes the resources of enterprises from the scope of the supply chain, optimizes the operation mode of modern enterprises, and reflects the market's requirements for enterprises to rationally allocate resources. It has a significant effect on improving the business process of the enterprise and enhancing the core competitiveness of the enterprise.

It can be said that the traditional inventory management model has been unable to adapt to the requirements of inventory management in the information age due to the lack of a supply chain concept [3]. With the help of the enterprise management information system (ERP), the information of inventory management and control is an effective way to solve the traditional inventory management problems. For this reason, more and more companies are buying various information software to improve their information, and ERP software is favored by these companies as software that provides a systematic solution. However, with the construction and application of ERP, the supply chain inventory management presents new changes and characteristics and also faces problems such as the poor operation of inventory management and control modules. If enterprises still manage inventory according to the existing inventory control model, there will undoubtedly be a mismatch with the ERP system, which will lead to the effectiveness of inventory management and control. Therefore, this thesis is aimed at analyzing the problems and causes of inventory management and human resource management in the ERP environment, to optimize the inventory management and control in the ERP environment and to propose specific solutions and corresponding development measures, to enable enterprises to achieve reasonable control of inventory and improve market competitiveness.

## 2. Relevant Works

Since 2000, the country puts forward the "artificial intelligence technology [4–6] to drive industrialization" of the new industrialization road, China's enterprise information construction pace has accelerated significantly setting off the application of the system climax, a lot of domestic enterprises are actively implementing the system now and thousands of enterprises use ERP systems, and ERP applications are also expanding, for example, human resource management [7]. In today's market environment, enterprises rely only on strong production capacity, and excellent products are far from enough. To survive and grow, enterprises must reduce various costs and increase profitability. Informationization of enterprise management has become a necessary condition for operators to make full use of information resources, grasp market opportunities, and better organize human, material, and financial resources for production and operation activities. Research on inventory management can be traced back to the 1930s and emerged in the 1980s in Western countries. The research on inventory management in China started in the 1990s, and due to the rapid development of China's economy, a hundred schools of thought have emerged in this area in the 21st century. This paper will collect the research literature on safety stock, ABC classification, sales forecasting, inventory management, and CMI from domestic and foreign experts and scholars [8].

The literature focuses on the safety stock setting of general-purpose materials, analyzes the factors affecting the safety stock quantity, and concludes that the procurement lead time is the main factor affecting the safety stock quantity, and the enterprise can optimize the procurement chain to reduce the procurement lead time, accelerate the material turnover speed, and finally achieve the purpose of reducing the safety stock. The literature focused on the safety stock setting of general-purpose materials, analyzed and studied the factors affecting the safety stock quantity, and concluded that the procurement lead time is the main factor affecting the safety stock quantity, and the enterprise can optimize the procurement chain to reduce the procurement lead time and accelerate the material turnover speed and finally achieve the purpose of reducing the safety stock. The literature studied how to set safety stock based on material requirement planning management. The premise of their study is how to carry out the most effective material requirement planning exercise in a given business environment to set a minimum safety stock level while ensuring production. They listed the factors that affect the setting of safety stock levels to form a mesh model and then analyzed each factor individually to determine the extent to which it affects whether the final order is delivered on time. It was eventually concluded that the relevant factors involved in safety stock affect each other but do not directly influence the final delivery outcome. The literature is based on the now popular supply chain management model for safety stock research as well as the ERP system to this classification of products and then the supply chain management strategy and safety stock management strategy for these categories of products, respectively, through a large amount of data; to come up

with different categories of products should be how to ensure the effective operation of the supply chain method under the limited amount of safety stock, so that the enterprise thus reduce inventory and achieve the purpose of reducing production costs [9].

Literature is combined with the practical situation of operation, using a qualitative forecasting method, based on the current market demand situation of the company; sales staff and customers were asked and surveyed, and then, statistical analysis was conducted based on the results of the survey, and finally, the demand forecast data was derived. Eventually, the actual order situation in the market confirms the validity of its demand forecast and provides a more reliable basis for the company's marketing decisions. The literature tried to attempt a quantitative analysis of demand forecasting with the help of a fuzzy linear regression function tool for demand forecasting. In the research process, they added qualitative demand forecasting analysis and achieved good results after combining qualitative and quantitative analyses for demand forecasting, thus verifying the feasibility of using a combination of qualitative and quantitative forecasting and providing valuable research ideas for subsequent researchers [10]. The literature used the GM(1, 1) grey model to analyze and forecast the future revenue trend of direct selling companies and predicted that the revenue of direct selling companies will rebound in the next quarter, using Amway, a US direct selling company, as a research sample. The main objective of the paper, which is specifically focused on inventory management in the medical device industry, is to verify which model is most effective in controlling inventory to achieve optimal cost and speed in the inventory control process. He conducted a pilot study using the inventory management model, the JIT model, and the zero inventory management model, respectively. In the end, the benefits of the inventory management model were significantly higher than those of the other two models in their particular circumstances [11].

The literature studies inventory in multilevel supply chains. They point out that in today's world of market activities, the division of labor is getting more and more detailed so that the supply chain lines are getting longer and longer, making their supply chain control quite difficult for companies under of the chain. If each point in the chain has its inventory management, it is quite passive at the end of the chain, while the distortion of data at each point is magnified and the risk of inefficient operation is increasing. If each point in the supply chain adopts an inventory management model, the efficiency of the whole supply chain will be improved exponentially, and the inventory of each point in the supply chain will be significantly reduced. Using the multiplier principle, it is clear that the more levels of the supply chain, the more significant the relative effect of the inventory management model [12].

### **3. ERP System Inventory Management and Human Resource Management**

*3.1. ERP System Operation Model.* Enterprise Resource Planning (ERP) is an information system, so the standard name

should be ERP. ERP takes information technology as the carrier and stores the data through the database, and the user operates through the front-end interface, injecting the enterprise management idea into it. The core of the ERP system is the management culture of the enterprise, through the input of data, computing, and then output data to provide decision-making tools for employees at all levels of the enterprise and the decision-making level [13].

For requirement planning, the core of MRP is object-centric; its focus is on material requirement planning, such as Figure 1. It is the ERP system framework hierarchy diagram. After the optimization of this system, the supplier's inventory management and human resource management can be well optimized, to achieve better utilization of the entire resources.

With the continuous development of enterprises and the introduction of supply chain ideas, the ERP system was born. The ERP system is based on the MRP system and constantly expands its functional modules and the management of enterprise ideas into each management module. From the user's point of view, the ERP system is divided into a data entry module, process approval module, and report module. The refinement of enterprise management, customer relationship management system, supply chain management system, personnel attendance system, product life cycle management system, etc., as shown in Figure 2, shows the ERP system optimization of the main module layout.

Safety stock, also known as insurance stock (or minimum stock, as some companies call it), exists because of unexpected events. For example, if a customer suddenly places an order and requires delivery outside the standard supply chain capacity (ignoring the company's production capacity for the moment), the company cannot meet the customer's demand without safety stock levels. It is the uncertainty of external factors (customers may place sudden purchase orders, and suppliers may not have the material available or may not be able to deliver the item on time) that makes safety stock so important. As mentioned earlier, resources are scarce for all businesses. Companies cannot unconditionally stockpile large quantities of raw materials to meet customer demand. Nor is it possible to produce large quantities of finished goods to respond to urgent customer demand. In terms of optimizing efficiency, zero inventory is, of course, the lowest cost state, but in the real business process, affected by the capacity of the supply chain, the production capacity of the enterprise, and the management level of the enterprise, the difficulty of achieving zero inventory is quite huge, and any problem in one of the links will affect the normal operation of the enterprise [14].

All the above reasons lead to the examination of safety stock management. It is relatively easy to calculate safety stock, but it is not easy to set a safety stock that suits the needs of the company. Therefore, every production-oriented company needs to set the right level of safety stock. A high level of safety stock can increase customer satisfaction and improve response time to customer needs, but it leads to the inventory taking up too much of the company's working capital. At the same time, if products are being replaced at a rapid rate, it also tends to cause some

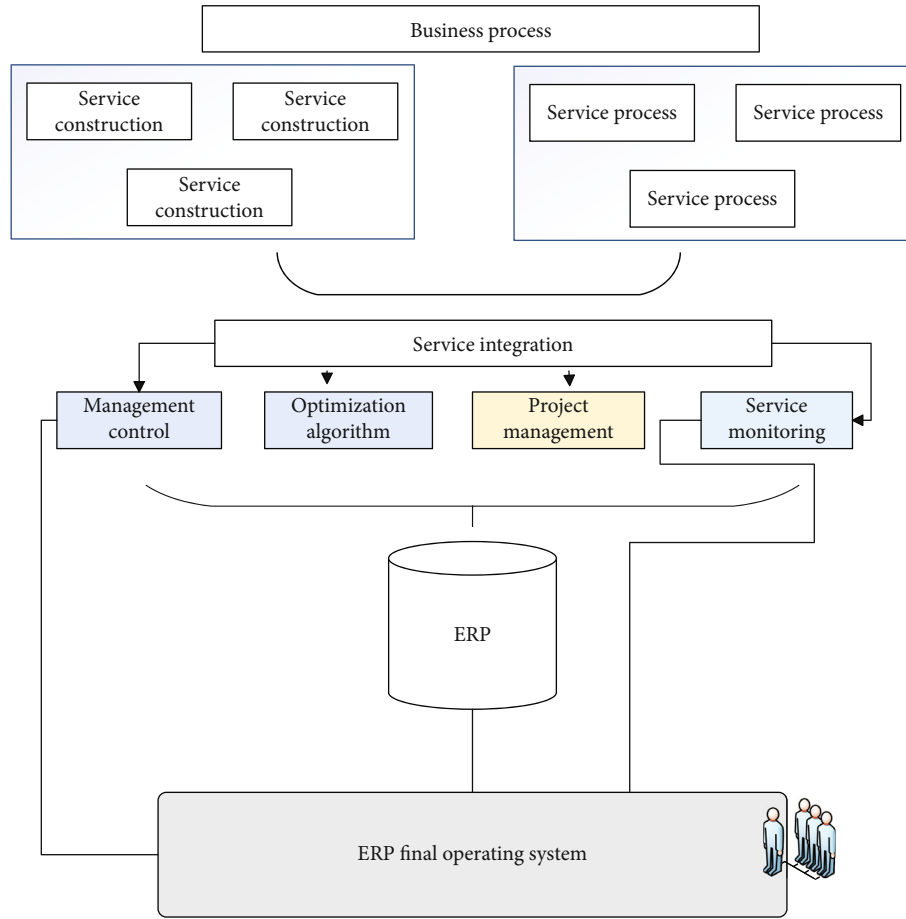


FIGURE 1: ERP system framework hierarchy diagram.

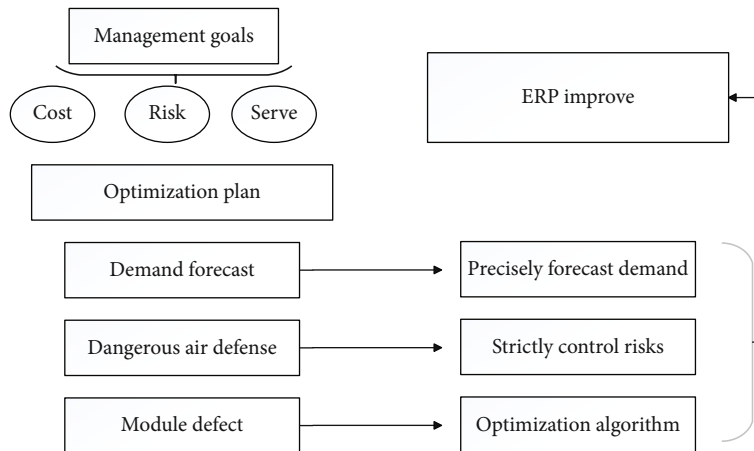


FIGURE 2: ERP system optimization main module layout.

nongeneric materials to become stagnant and may eventually be simply scrapped, leading directly to an unproductive drain on the company’s working capital. On the other hand, if the safety stock is too small, although it can reduce the backlog of the enterprise’s working capital, when there are unexpected events, the enterprise cannot effectively respond and ultimately affects the competitiveness of the enterprise.

Therefore, the right safety stock is a balance between cost and service level, which can meet the customer’s delivery needs in general and also the inventory cost and inventory risk of the enterprise. The right safety stock is different for each company because three factors affect the safety stock setting; they are the procurement lead time of materials, the company’s service level, and the variation of customer

demand; each company has a different focus and supply chain management level, so the final safety stock quantity will also vary.

The SS formula for a safety stock consisting of three elements is

$$S(j) = \frac{1}{n} \sum_{i=1}^n X_i Y_i. \quad (1)$$

**3.2. Inventory Management Process in ERP.** The implementation of the entire supplier inventory management is transparent, and the buyer's enterprise and the supplier can monitor it at any time. It is mainly divided into two parts: Inventory management: in fact, it is composed of sales forecasting and inventory management as well as the supplier's production system because after the supplier's inventory management is implemented, the work of these parts is mainly completed by the coordination of the supplier and the buyer's enterprise. So, it is classified as a module to deal with: first, the buyer's company obtains the sales data of the product and then combines it with the current inventory level and transmits it to the supplier in time, and then, the supplier's inventory management system makes a decision: if the supplier or the existing warehousing system can meet the number of products required by the inventory management system to make decisions, the warehousing and transportation distribution system will directly deliver the products to the buyer's enterprise promptly. If the supplier's existing warehousing system cannot meet the inventory management system, to make a decision, it is necessary to notify the production system to produce the product and then deliver the product to the buyer's enterprise promptly through the transportation and distribution system. Among them, before the formal order is generated, it should also be handed over to the buyer's enterprise for verification, and the final order can be obtained after adjustment.

Supplier managed inventory is a new inventory management model developed to reduce costs for both sides of the transaction. Inventory management is based on the rapid response and effective customer response management model. The core idea of "sharing" is that both supply and demand sides share their effective inventory data and current and future demand data and then replenish the stock according to the actual purchase or consumption data of the demand side. Due to the sharing of information, the supply and demand sides save unnecessary communication costs and unnecessary guesswork, avoid unnecessary inventory backlogs, and thus reduce the total cost of the entire supply chain. Under the traditional inventory management model, each enterprise in the supply chain is working on its own, playing the internal calculations of the enterprise, and not managing inventory from an integrated and holistic view. The emergence of inventory management breaks the old model of inventory management, in which both supply and demand sides place objects in the physical area of the demand side or a third party based on a common agreement, and the supply side manages the inventory. The agreement is optimized and adjusted based on the experience of both parties so that the

supply and demand sides can continue to work together better. From the point of view of property rights, although the supplying party's items are delivered to the area designated by the demander, the property rights still belong to the supplying party and there is no liability on the demander's side, but the demander needs to take certain responsibility for the storage of the items. Only when the demander takes possession of the items in the inventory management warehouse do the property rights and claims change. Figure 3 shows the flow chart of the production and inventory business of the company based on the ERP system [15].

The implementation of an inventory management model is beneficial to both the supply and demand sides. For the supply side, inventory management can stabilize customer relationships while reducing transportation costs and improving delivery efficiency; for the demand side, it not only reduces the pressure of inventory on working capital but also reduces the risk of material shortages and speeds up the supply chain response. With the widespread use of inventory management, the initial weakness of the supply side is gradually reversed and the related inventory management agreements are gradually improved. For inventory management, the setting and management of inventory levels are also more scientific, as the maximum and minimum values of inventory are set based on past transaction data and future demand forecasts of both parties to determine specific inventory targets. The inventory management model has been applied by domestic and foreign enterprises for more than 20 years, especially in large enterprises such as Wal-Mart, P&G, Amazon, Foxconn, and Heller. The inventory management model has been proved to be an effective and advanced inventory management model with a lot of advantages, which can be summarized as follows: (1) During the agreement cycle, there is no transfer of property rights and no claims occur. Therefore, the demand side does not experience the situation of inventory crowding out the operating capital of the enterprise, which can reduce its operating costs for the demand side and directly enhance its market competitiveness. (2) The sharing of information between the supply and demand sides helps to reduce the communication cost and improve the communication efficiency of both sides and also reduces the "bullwhip effect" in the demand forecast between the supply and demand sides. (3) Due to the zero response time of object supply, the productivity of the demand side can be improved, thus improving its customer response time and customer satisfaction. (4) It helps the supplier to stabilize its customer relationship and increase its customer stickiness. The inventory management model has a lot of advantages but also has a lot of limitations, which can be summarized in the following aspects. (1) Since the core of the inventory management model is information sharing, it is highly dependent on the information technology of the enterprise. (2) In the inventory management model, the demand side is obviously in a dominant position, and if the underwriting terms are not agreed upon through coordination, the cooperation between the two parties is still a "zero-sum" game. (3) The inventory management model requires a certain level of financial and managerial strength on the part of the supplier. In the absence of an

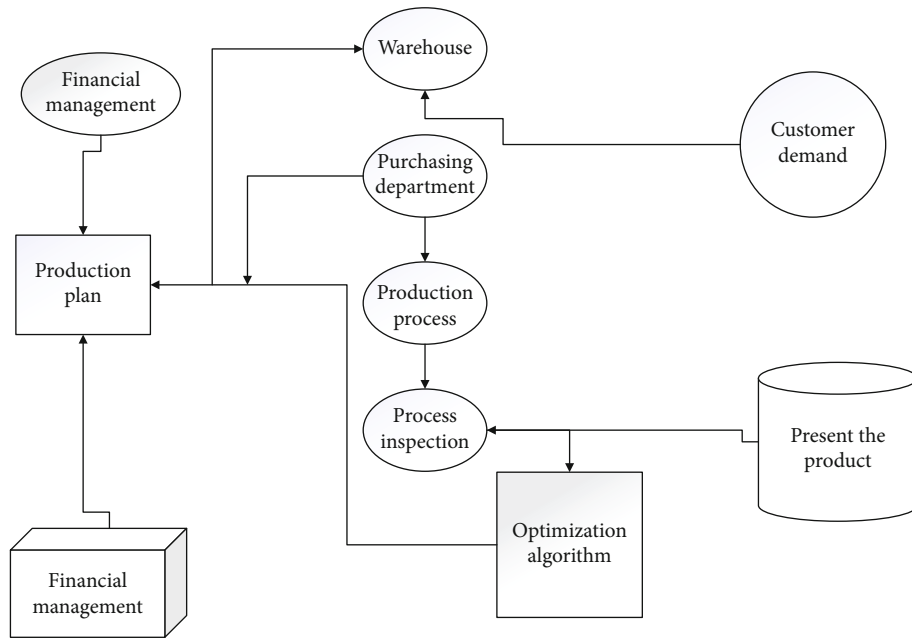


FIGURE 3: Flow chart of the company's production and inventory business based on ERP system.

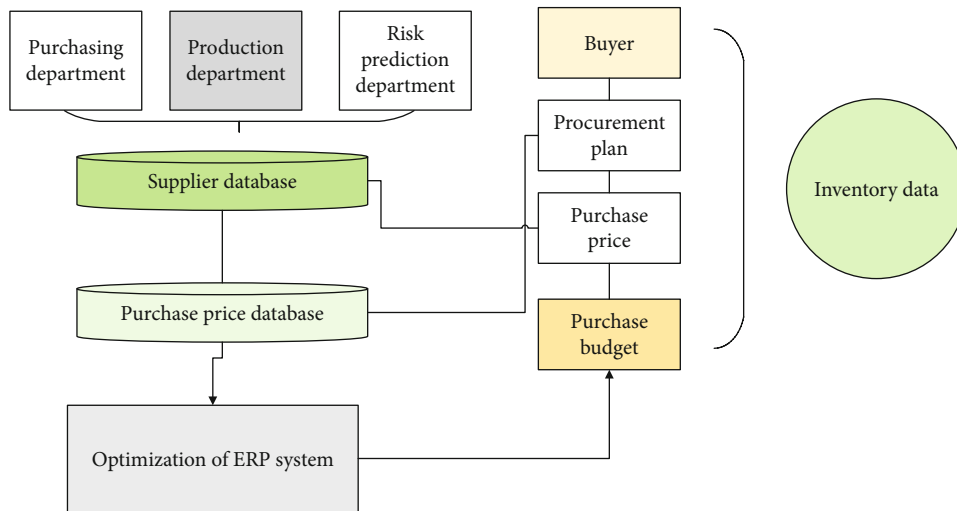


FIGURE 4: Flow chart of the company's ERP-based procurement and payment process.

equal underwriting agreement [16], the supplying party is exposed to the risk of inventory build-up. Figure 4 shows the procurement and payment process of the ERP system.

3.3. *Human Resource Management Strategies in ERP.* The job description is an important international tool for the company's human resource development and management. It has a fundamental role in improving work performance, performance evaluation, job training, standardizing management processes, and goal management. A considerable proportion of companies in our country are not accustomed to using job descriptions for human resource development and management, which is extremely detrimental to the company's development. It is expected that more and more companies will introduce and use "job descriptions" to optimize

the allocation of human resources and improve the level of human resource management.

In today's increasingly fierce business competition between enterprises, how to attract the best talent, rationalize human resources, reduce personnel costs, and improve the competitiveness of enterprises have been the primary consideration of enterprise managers; that is, to consider the addition of ERP in the human resources system has made its functions expanded to a full range of enterprise management. The scope of HR functions has also developed from a single payroll accounting and personnel management to a full range of solutions that can help the enterprise's decision-making. These areas include human resource planning, employee evaluation, workforce scheduling, time management, recruitment management, employee payroll,

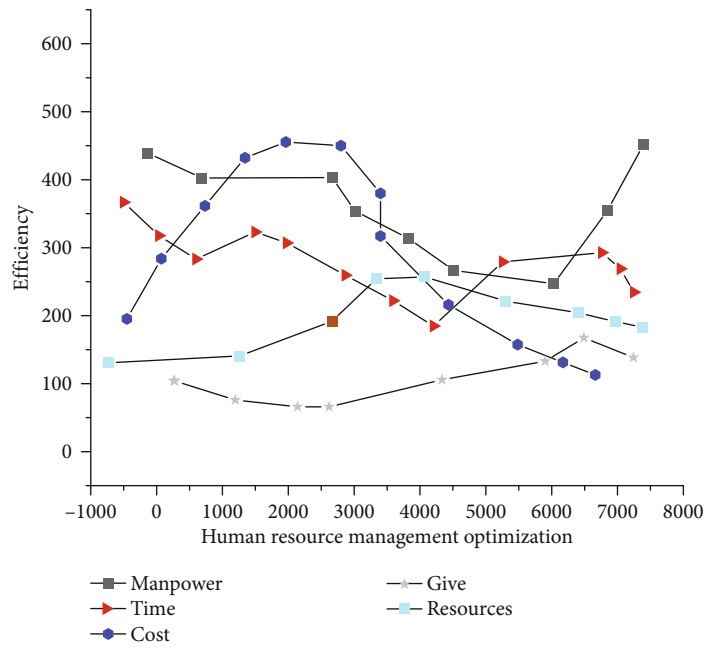


FIGURE 5: Efficiency of traditional HRM in ERP.

training programs, and travel management. Together with the financial and production systems in ERP, they form an efficient and highly integrated enterprise resource system. Figure 5 shows the efficiency of the traditional HR management compared with the optimized one [17].

The HR system is used to support decision-making in human resources planning. Managers can use the HR system in ERP to easily develop their own organizational and staffing planning plans based on their production needs. The comparison of various plans in the system and the evaluation of simulated operations generate result data for various plans and provide support for managers' final decisions through an intuitive graphical user interface. Human resource planning also allows the development of job models including job requirements, promotion paths and training plans, and a series of training recommendations for the employee based on the qualifications and conditions for the position. In addition, personnel cost analysis allows for analysis and forecasting of past, present, and future personnel costs and provides the basis for corporate cost analysis through an integrated ERP environment [18].

*Recruitment management.* A company's workforce should be seen as the most important investment. Human resource decisions are about the success and competitiveness of a company. Remaining competitive means that a company has a set of effective tools to identify talent. People are a company's most important resource. Good people are the only way to ensure a company's lasting competitiveness. Recruitment systems generally support the recruitment process in several ways: they optimize the recruitment process and reduce operational workload; they reduce recruitment costs by scientifically managing the cost of recruitment; they provide information to support the selection of hiring positions and effectively help the company in its talent search [19].

*Payroll accounting.* The flexible and efficient payroll system can develop payroll accounting methods according to the different payroll structures and processing flows of the company across regions, departments, and jobs. Direct integration with time management allows for timely updates to employee payroll dynamics. Manual intervention is reduced, and problems in the interface are eliminated, providing automatic payroll deductions, employee loans, and other functions. Payroll management can get the required information in advance through payroll simulation runs. The payroll system also has a powerful recalculation function. When the payroll accounting process is completed, the master data of the employee concerning the previous payroll period changes, and the recalculation function is automatically triggered in the next payroll accounting period for correction. In addition, the system can also automatically adjust the salary structure according to the results of employee assessment [20].

*Job management.* Job management is based on the national or local calendar, flexible scheduling of business operations and workforce schedules, and a comprehensive set of arrangements for employee overtime, work shifts, employee holidays, and employee relief. With the remote attendance system, the actual attendance status of employees can be recorded in the main system. Time data related to employee payroll and bonuses are further processed in the payroll system and cost. The system includes time management as an integral part of the overall system, and this system supports the planning, control, and management processes of the HRM system. Figure 6 shows the optimization rate of the human resource management system.

The ERP system can automatically control the entire process from travel application travel approval to travel reimbursement for workflow control. The entire process



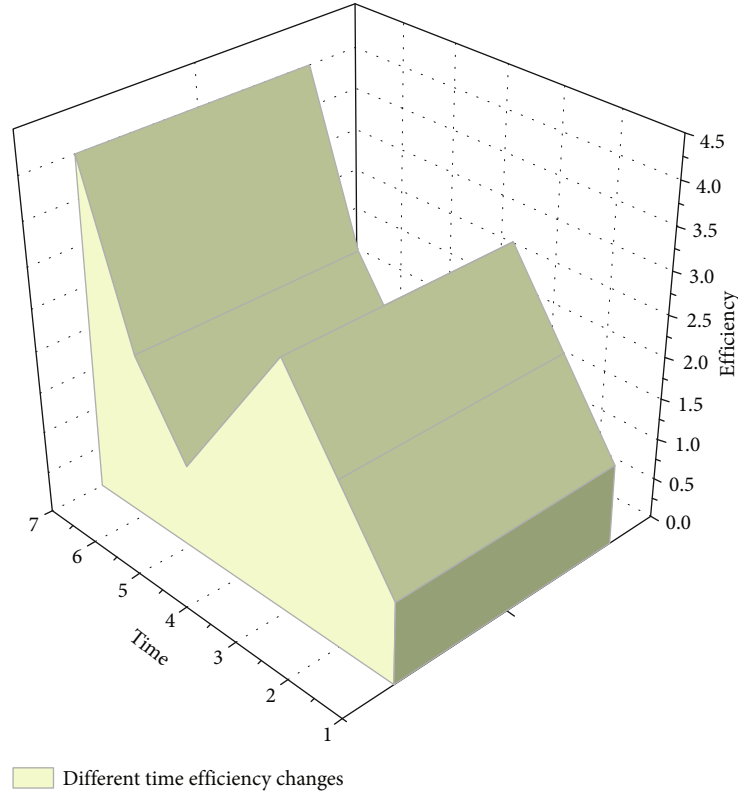


FIGURE 6: Human resource management system optimization rate.

can be completed in the system and through the integrated environment will be accounting data into the financial costing module to go. The previous ERP system was centered on the manufacturing and sales process (supply chain). Therefore, for a long time, the resources related to manufacturing resources have been managed as the core resources of the enterprise. However, in recent years, human resources within the company have become more and more important to the company and are considered to be the foundation of the company's resources. In this context, human resource management has been added to the ERP system as an independent module, and the financial and production systems in ERP form an efficient and highly integrated enterprise resource system [21].

#### 4. Research and Development of ERP Systems

**4.1. Study of ERP Systems.** The application of ERP reflects the idea of "lean production and agile manufacturing" in supply chain management, allocate resources, and quickly meet market demand. The supply chain is the core competitiveness of modern enterprises; the ERP system achieves the enterprise supply chain refinement and collaborative management and improves the competitiveness of enterprises. The ERP system has a fine planning system, including the main production plan, material demand planning, procurement planning, sales planning, financial budget, and human resource planning. These planning and value control functions are used in the enterprise's entire supply chain man-

agement for the enterprise production and operation with "prior control" conditions. The formula for calculating time is the same time:

$$\text{Time} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2. \quad (2)$$

ERP can synchronize business information with financial information in real time; enterprise management can analyze data and financial analysis functions through ERP system, real-time monitoring of procurement, production, sales, and traceability of business activities, for the enterprise's production and operation and human resource management for "control in the matter," and to make timely decisions. For timely decision-making, Table 1 is the comparison of the relationship between time and distance in inventory management under ERP [22].

ERP applications are customizable and flexible, and at the early stage of ERP system design, enterprises can select the functional modules that need to be implemented according to their business characteristics and management control needs and personalize the business processing and flow of the ERP system to effectively support the daily operation and management of the enterprise. In the process of ERP system application, with the business development and management needs, enterprises can adjust the ERP system environment by resetting system parameters, system upgrades, etc., and if necessary, secondary development of

TABLE 1: Inventory management model under ERP model.

Distance/time	1	2	3	4	5	6	7	8	9	10
10	5	6	7	8	9	10	11	12	13	14
20	5	7	9	11	13	15	17	19	21	23
30	7	8	7	7	7	67	4	2	7	7
40	5	6	7	8	9	10	11	12	13	14
50	8	55	66	3	45	66	34	53	56	21
60	5	66	4	7	10	13	16	19	22	25
70	7	67	12	34	56	44	45	56	45	45
80	5	53	15	24	43	32	43	56	43	21
90	6	23	18	12	34	56	34	56	43	23

the system program to strengthen system functions and improve the integration and refinement of the system to meet the changing needs of business management [23].

ERP uses the idea of planning and control to process, standardize, and refine the management of enterprise operation. In the ERP environment, enterprises can achieve comprehensive control of people, finance, and materials in all aspects such as production, supply, and sales and realize the integration of financial and business processing. The following content analyzes the characteristics of internal control in an ERP environment from the aspects of five elements of internal control, such as control environment, risk assessment, control activities, information and communication, and internal supervision. Table 2 is the comparative data of the respective algorithm functions.

In an ERP environment, the organizational structure of the enterprise is flattened and the quality of the personnel is required to be higher. In the traditional environment, the enterprise internal pyramid structure, through the different levels and functions of each department set up for the daily work of the enterprise to allocate and control, each department's business processing is relatively independent. In the ERP environment, the operation of the enterprise from the departmental functions as the center to the business process as the center is according to the business module to carry out. For example, ERP from procurement to payment, from order acquisition to capital recovery, and other businesses are fully integrated; a variety of business activities are closely integrated with the logistics and capital flow, and information flows together, for the internal transfer of information across functional departments to provide an integrated platform to achieve cross-departmental business processing and information sharing, breaking the original functional departmental compartmentalization and hierarchical division of a clear organizational structure. In the ERP environment, the internal control hierarchy is significantly reduced and the control responsibilities are clearer, which is conducive to the improvement of control efficiency. At the same time, the ERP system can improve the office efficiency of enterprises, so that enterprise managers and employees can use the information system operation to play their part of the function, using system operations instead of manual operations. The human-machine cooperative work mode makes the employees from the complicated manual operation to free, so the enterprise functions

TABLE 2: Efficiency test results of ERP optimization algorithm.

Algorithm/test function	1	2	3	4	5
ISPO	6.45	3.45	3.23	5.67	6.57
TXRP	4.34	4.54	1.23	6.78	5.67
HM-PSO	4.52	3.45	2.34	3.45	6.76
WEIFN	2.34	3.45	3.23	3.54	5.45

should pay more attention to the management function rather than simple transaction processing. In addition, the professionalism and complexity of ERP system operations on the quality of enterprise employees put forward higher requirements: information system management departments need professional and technical personnel to carry out maintenance and management of the system to ensure the safe and stable operation of the system; end-users need to improve the understanding of ERP management ideas and the ability to operate the system based on the original business skills. As the ERP system needs to be upgraded continuously with the changes in business management, employees need to learn continuously to adapt to the changes in system updates. Therefore, the implementation of ERP has put forward higher requirements on the overall quality of enterprise employees, so the human resource management should be more careful [24].

*4.2. Optimal Development of the ERP System.* This section proposes the content and methods of optimizing enterprise internal control in ERP environment; that is, enterprises should make full use of ERP system functions to optimize internal control measures and analyze the potential risks of each link in business processes from existing business processes, improve relevant management systems, and implement key control points by formulating or designing corresponding manual or automated control measures, to achieve internal control objectives. In addition, enterprises in the ERP environment need to strengthen the internal control of information systems; information system internal control includes general control and application control because the information system application control has been built-in in the system development, so the information system in the daily business management activities of enterprises focused on strengthening the general control of the information system management department, as shown in Figure 7 of the ERP system development flow chart.

In the ERP environment, enterprises should make full use of the ERP system functions for the corresponding development, specifically analyzed as follows.

*4.2.1. Optimizing Incompatible Job Separation and Authorization Control Using System Parameters and Permission Settings.* Enterprises implementing ERP need to analyze and sort out the incompatible functions in the business process of the system and set up different user logins for different functions, and the interface that each user can operate after logging into the system is different, and the functional scope of the interface is different. For example, the accountant who is responsible for bookkeeping can only perform the bill-making operation after logging into the

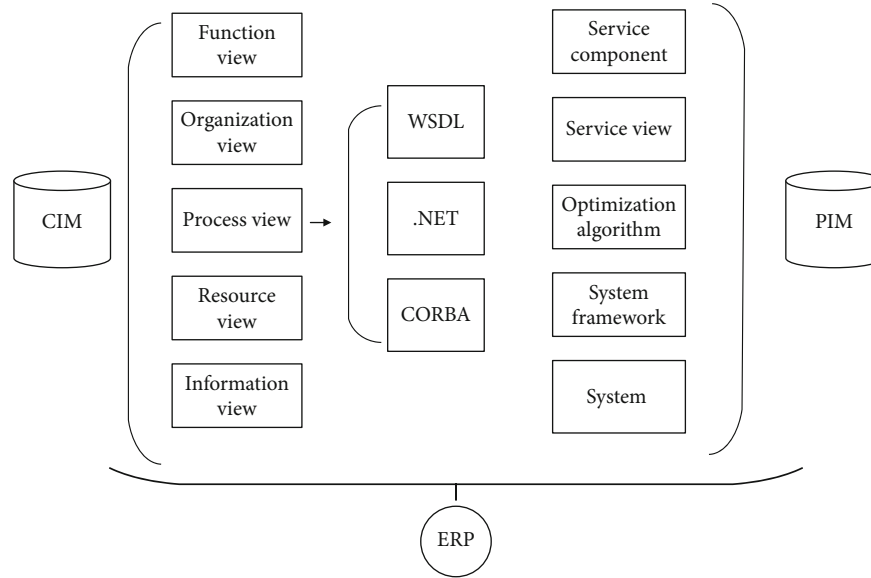


FIGURE 7: ERP system development flow chart.

system and upload the bookkeeping vouchers to the financial controller, who will review the bookkeeping vouchers in his operation interface after logging into the system, and the system does not allow the bill maker and the reviewer to be the same person. Therefore, the implementation of ERP enterprises should be in the full and reasonable use of the ERP system permissions set for the separation of incompatible function control. In addition, ERP for user authorization is through the role or permission parameter file with different permissions granted to different users, after the authorization of the user master data recorded in its authorization information. When a user executes a transaction code, the program corresponding to the transaction code checks the authorization information in the user's master record for permissions, and if the check passes, it means that the user has the operation authority, and the business operation can be carried out smoothly; otherwise, the system automatically terminates the transaction processing and prompts the user to operate without authority. The system's parameterized control of each position and function enables system users to clarify their respective terms of reference and strengthens the control of the corresponding rights and responsibilities of each position through system control.

**4.2.2. Optimize Approval Control Using System Approval Process Settings.** Enterprises need to embed the approval process of each business link in the ERP system and set the user with approval authority at the approval node, the user logs into the system and carries out the approval operation for the approval transaction to be processed, the system will pass the data to the next link of the approval process according to the approval result, and the corresponding approval data will be saved in the database. The approval function in the ERP system can standardize the approval process of enterprise business processing and improve approval efficiency, which is a powerful tool for enterprise approval control.

**4.2.3. Optimization of Accounting System Control Using Financial Management Module.** The formula is

$$f(x) = \frac{1}{n} \sum_{i=1}^n X_i^2 \left( \frac{x - \mu}{\sigma} \right). \quad (3)$$

The financial management module can automatically collect the original data and production data generated from sales, purchase, and production activities and use the corresponding data to generate general ledger and accounting reports, avoiding multiple entries of the same data and reducing errors caused by human operations. The organizational structure, accounting system, financial standards, and rules of the enterprise need to be defined in the general ledger module, and all financial-related business information generated by submodules will be passed to the general ledger, which will be processed to generate general ledger data to reflect the financial results of the enterprise. The formula for calculating efficiency is

$$E = \frac{1}{n} \sum_{i=2}^n X_i^2 \frac{x - \mu}{\sigma}. \quad (4)$$

The full and reasonable use of ERP financial management module functions can help enterprises to expand the scope of accounting supervision and control from local to the whole process of business activities. In addition, the accounting information generated by the ERP system is multidimensional and real time, so that the accounting supervision and control functions can be given full play. Enterprises can use ERP systems to extend the scope of accounting supervision and control from the local to the whole process of business activities.

**4.2.4. Optimize Property Protection Controls Using Fixed Asset Management and Inventory Management Modules.** Fixed asset management in the ERP system is based on fixed asset card management, which helps enterprises to realize

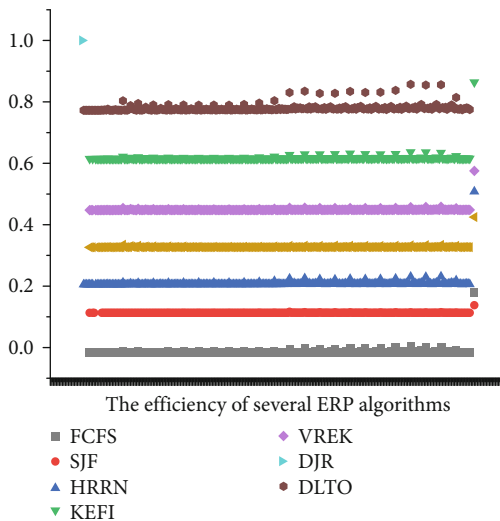


FIGURE 8: The efficiency of ERP system after optimization.

the comprehensive management of fixed assets. The fixed asset management function includes the business processing of fixed asset addition, reduction, original value replacement, asset evaluation, departmental transfer, internal transfer, status change, etc. It also allows the operation of depreciation and provision for impairment following the requirements of national accounting standards. At the same time, the fixed asset management function can help managers fully grasp the current quantity and value of fixed assets, track the use of fixed assets, and strengthen enterprise asset management. The inventory management module can handle the inventory management business of purchasing, production, sales, and other departments and establish inventory files for all materials. Inventory materials can exist in a variety of inventory states, such as regular inventory, vendor-hosted inventory, and in-transit inventory. Different inventory states can be managed separately, and the inbound, outbound, movement, and inventory of inventory materials can be fully controlled and managed. After the implementation of ERP, enterprises should regularly maintain the fixed asset management and inventory management modules of the system to make up for the problems that occur in the operation of the system and optimize and upgrade the system when necessary to make full use of the ERP system functions and strengthen property protection control. Figure 8 shows the results of the efficiency of the ERP system after the optimization is achieved.

## 5. Conclusion

After the system is optimized, the processing efficiency is higher, and the inventory can be optimized to promote resource conservation. The purpose of building an information management platform for enterprises is to improve long-term management concepts, establish an optimized resource management system, better optimize the allocation of enterprise resources, and enhance core management capabilities. The multifunctional data analysis capability of the ERP system is used to realize the optimization of budget

management in financial management, improve the enterprise budget management system, and better achieve the long-term strategic goals of the enterprise. This paper provides an in-depth overview and analysis of the budget management system of a representative manufacturing enterprise through the perspective of the integration of the ERP system and human resource management and proposes improvement suggestions for the defects and problems that occur in the process of its application, to optimize the resource allocation of the information platform of the enterprise, and proposes how to make the system achieve the management of inventory and the optimization of human resource management. In this way, the enterprise can use the ERP system to achieve optimal benefits, both for inventory management and the management of enterprise manpower. For the application of ERP, I want to enable the entire industry to better optimize the use, so that merchants' inventory processing can be more optimized, and in the future use, the process can reduce losses and promote better development of the industry.

## Data Availability

The data used to support the findings of this study are included within the article.

## Conflicts of Interest

All the authors do not have any possible conflicts of interest.

## References

- [1] R. van Erp, R. Soleimanzadeh, L. Nela, G. Kampitsis, and E. Matioli, "Co-designing electronics with microfluidics for more sustainable cooling," *Nature*, vol. 585, no. 7824, pp. 211–216, 2020.
- [2] D. Kondziella, A. Bender, K. Diserens et al., "European Academy of Neurology guideline on the diagnosis of coma and other disorders of consciousness," *European Journal of Neurology*, vol. 27, no. 5, pp. 741–756, 2020.
- [3] C. Foster, D. Boulosa, M. McGuigan et al., "25 years of session rating of perceived exertion: historical perspective and development," *International Journal of Sports Physiology and Performance*, vol. 16, no. 5, pp. 612–621, 2021.
- [4] M. Zhao, A. Jha, Q. Liu et al., "Faster mean-shift: GPU-accelerated clustering for cosine embedding-based cell segmentation and tracking," *Medical Image Analysis*, vol. 71, p. 102048, 2021.
- [5] Y. Gu, A. Chen, X. Zhang, C. Fan, K. Li, and J. Shen, "Deep learning based cell classification in imaging flow cytometer," *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 2, pp. 18–27, 2021.
- [6] M. Zhao, Q. Liu, A. Jha et al., "VoxelEmbed: 3D instance segmentation and tracking with voxel embedding based deep learning," 2021, <http://arxiv.org/abs/2106.11480>.
- [7] T. Van Erp and D. Sanders, "Demands of professional cycling races: influence of race category and result," *European Journal of Sport Science*, vol. 21, no. 5, pp. 666–677, 2021.
- [8] M. S. E. Sendi, G. D. Pearlson, D. H. Mathalon et al., "Multiple overlapping dynamic patterns of the visual sensory network in

- schizophrenia,” *Schizophrenia Research*, vol. 228, pp. 103–111, 2021.
- [9] H. Arnts, W. S. van Erp, J. C. M. Lavrijsen, S. van Gaal, H. J. Groenewegen, and P. van den Munckhof, “On the pathophysiology and treatment of akinetic mutism,” *Neuroscience & Behavioral Reviews*, vol. 112, pp. 270–278, 2020.
- [10] G. E. Benoist, I. M. van Oort, E. Boerrigter et al., “Prognostic value of novel liquid biomarkers in patients with metastatic castration-resistant prostate cancer treated with enzalutamide: a prospective observational study,” *Clinical Chemistry*, vol. 66, no. 6, pp. 842–851, 2020.
- [11] K. Westerdijk, I. M. E. Desar, N. Steeghs, W. T. A. Graaf, N. P. Erp, and on behalf of the Dutch Pharmacology and Oncology Group (DPOG), “Imatinib, sunitinib and pazopanib: from flat-fixed dosing towards a pharmacokinetically guided personalized dose,” *British Journal of Clinical Pharmacology*, vol. 86, no. 2, pp. 258–273, 2020.
- [12] E. Boerrigter, L. N. Groen, N. P. van Erp, G. W. Verhaegh, and J. A. Schalken, “Clinical utility of emerging biomarkers in prostate cancer liquid biopsies,” *Expert Review of Molecular Diagnostics*, vol. 20, no. 2, pp. 219–230, 2020.
- [13] W. Woliner–van der Weg, M. Peppelman, Y. S. Elshot et al., “Biopsy outperforms reflectance confocal microscopy in diagnosing and subtyping basal cell carcinoma: results and experiences from a randomized controlled multicentre trial\*,” *British Journal of Dermatology*, vol. 184, no. 4, pp. 663–671, 2021.
- [14] T. C. Zwart, D. J. A. R. Moes, P. J. M. van der Boog et al., “Model-informed precision dosing of everolimus: external validation in adult renal transplant recipients,” *Clinical Pharmacokinetics*, vol. 60, no. 2, pp. 191–203, 2021.
- [15] S. Mikdad, I. A. M. van Erp, M. El Moheb et al., “Pre-peritoneal pelvic packing for early hemorrhage control reduces mortality compared to resuscitative endovascular balloon occlusion of the aorta in severe blunt pelvic trauma patients: a nationwide analysis,” *Injury*, vol. 51, no. 8, pp. 1834–1839, 2020.
- [16] P. A. Everts, A. van Erp, A. DeSimone, D. S. Cohen, and R. D. Gardner, “Platelet rich plasma in orthopedic surgical medicine,” *Platelets*, vol. 32, no. 2, pp. 163–174, 2021.
- [17] J. B. Carnevale and I. Hatak, “Employee adjustment and well-being in the era of COVID-19: implications for human resource management,” *Journal of Business Research*, vol. 116, pp. 183–187, 2020.
- [18] H. Opatha, “The coronavirus and the employees: a study from the point of human resource management,” *Sri Lankan Journal of Human Resource Management*, vol. 10, no. 1, pp. 37–49, 2020.
- [19] I. Gigauri, “Influence of Covid-19 crisis on human resource management and companies’ response: the expert study,” *International Journal of Management Science and Business Administration*, vol. 6, no. 6, pp. 15–24, 2020.
- [20] P. Boselie, J. Van Harten, and M. Veld, “A human resource management review on public management and public administration research: stop right there...before we go any further...,” *Public Management Review*, vol. 23, no. 4, pp. 483–500, 2021.
- [21] G. Dhiman, K. K. Singh, A. Slowik et al., “EMoSQA: a new evolutionary multi-objective seagull optimization algorithm for global optimization,” *International Journal of Machine Learning and Cybernetics*, vol. 12, no. 2, pp. 571–596, 2021.
- [22] L. Abualigah, “Group search optimizer: a nature-inspired meta-heuristic optimization algorithm with its results, variants, and applications,” *Neural Computing and Applications*, vol. 33, no. 7, pp. 2949–2972, 2021.
- [23] J. S. Pan, N. Liu, S. C. Chu, and T. Lai, “An efficient surrogate-assisted hybrid optimization algorithm for expensive optimization problems,” *Information Sciences*, vol. 561, pp. 304–325, 2021.
- [24] N. Zeng, D. Song, H. Li, Y. You, Y. Liu, and F. E. Alsaadi, “A competitive mechanism integrated multi-objective whale optimization algorithm with differential evolution,” *Neurocomputing*, vol. 432, pp. 170–182, 2021.

## Retraction

# Retracted: Influencing Factors of Microlecture on the Teaching Effect of Ideological and Political Courses in Colleges

### Wireless Communications and Mobile Computing

Received 8 August 2023; Accepted 8 August 2023; Published 9 August 2023

Copyright © 2023 Wireless Communications and Mobile Computing. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

### References

- [1] Y. Jin, "Influencing Factors of Microlecture on the Teaching Effect of Ideological and Political Courses in Colleges," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 6309221, 7 pages, 2021.

## Research Article

# Influencing Factors of Microlecture on the Teaching Effect of Ideological and Political Courses in Colleges

Yuqian Jin <sup>1,2</sup>

<sup>1</sup>*School of Marxism, Northwestern Polytechnical University, Xi'an 710072, China*

<sup>2</sup>*Office of Faculty Development Center of Xi'an Technological University, Xi'an 710021, China*

Correspondence should be addressed to Yuqian Jin; [sdyuqian@126.com](mailto:sdyuqian@126.com)

Received 5 August 2021; Revised 23 August 2021; Accepted 24 August 2021; Published 30 September 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Yuqian Jin. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Microlecture has the characteristics of single topic, easy to learn, convenient sharing, and real-time interaction. Whether these characteristics are conducive to enhancing the effect of ideological and political teaching in colleges is the focus of this paper. We have constructed the influence factor model of microlecture on ideological and political teaching effect in colleges. Through questionnaire survey and empirical analysis, we verify the four characteristics of microlecture. The results show that in microlecture teaching, single topic, easy to learn, and sharing convenience are the main factors to enhance the teaching effect of ideological and political course in colleges, while real-time interaction had no significant effect. Our study has enriched the literature of microlecture in ideological and political teaching and confirmed that microlecture can improve the teaching effect of ideological and political courses in colleges. Based on the research results, we propose the following recommendations: (1) in ideological and political teaching, recorded microlecture should concentrate on a single topic as far as possible, and the time of microlecture should be controlled to be shorter; (2) since students believe that microlecture make their study easier and share convenience, universities should consider promoting this mode in ideological and political teaching.

## 1. Introduction

Along with the practice and application of flipped classrooms in colleges and universities [1, 2], a more novel teaching method, namely, microlecture [3], appears in students' daily learning. Generally speaking, microlecture refers to the use of a short recording or video (usually within 3 minutes) to state a single, strictly defined topic [4]. Microlecture usually consists of lectures or presentations, narration slides, or screen playback with narration [5]. Due to the characteristics of the microlecture, such as short, clear themes, and adaptable to network transmission, it can be easily shared on WeChat, Weibo, short videos, or YouTube [3].

Ideological and political courses [6] are a compulsory course for college students. Due to the needs of teaching innovation, the flipped classroom model has been gradually introduced into the course [7]. Adhering to the concept of autonomous learning in the inverted classroom, the microlecture teaching method has been widely used and promoted in ideological and political courses. The application of

microlecture in ideological and political courses (IaPC) has many advantages [8]. For example, the explanation of a single topic can make it easier for students to understand. The microvideo method is conducive to the dissemination of knowledge itself, and students can also easily obtain it through the Internet. Students can also post their own comments under the video to achieve the purpose of interaction. In addition, college teachers make full use of microlecture to obtain better teaching effects, improve students' enthusiasm and sense of happiness in learning ideological and political theory courses, and thereby improve teaching effects [9].

It is undeniable that microlecture have played a significant role in innovative ideological and political teaching effect. However, the specific factors used by microlecture to influence the effect of IaPC in colleges need to be further explored. For example, Zeng [10] proposed that as a new education model emerging at home and abroad, microlecture needs to introduce ideological and political education in colleges. However, the feasibility and effectiveness of the design of the teaching mode of ideological and political

education in colleges based on microlecture need to be studied in depth. Wu [11] believed that microlecture, as a new teaching method under the background of “Internet plus,” provides rich resources for ideological and political education and can expand its education direction. However, the neural network [12–15] information technology [16, 17] represented by the microlecture affects people’s thinking mode, value orientation, and development trend in the field of culture and thinking. But how does microlecture influence ideological and political teaching at colleges? This paper is centered on this question.

In view of this, according to the specific characteristics of the microlecture, this article constructs a model of the influencing factors of the teaching effect of IaPC in colleges. Our research data are derived from a survey of 237 students who have taken ideological and political courses, and SmartPLS software is used to verify the research hypothesis. The purpose is to explore the concrete influence of microlecture on the teaching effect of IaPC in colleges. Our study is of great significance to guide university teachers to improve ideological and political teaching and enhance teaching effect. Moreover, we have enriched the research of microlecture in the field of ideological and political courses.

The rest of the paper is arranged as follows: the second part is a literature review; the third part constructs the research model and puts forward research hypotheses; the fourth part introduces methods and provides empirical results; the fifth and final part presents the conclusions.

## 2. Literature Review

*2.1. Features of Microlecture.* As a new type of online learning resources, microlecture has been developed rapidly all over the world [18]. With the progress of information and communication technology and the widespread use of mobile smart phones [19], microlecture has also been mobilized from the original web page format.

According to the definition of microlecture [3–5] and practice [9–11], it can be found that microlecture has the characteristics of single content, easy learning, convenient sharing, and real-time interaction. Single content means that in the process of making a microlecture, teachers will generally explain a topic in a microlecture, so that the knowledge of the microlecture is focused [20]. Ease of learning means that students can easily learn and understand the video content due to the simplification of the subject of the microlecture [21]. Generally speaking, the subject content of a single microlecture is relatively small. There are many reasons for sharing convenience. First of all, the content of microlecture is small, and the speed of transmission is fast; secondly, the wide application of mobile smart devices in the crowd provides a huge user base for microlecture [22]. Interaction refers that teachers and students, students and students can comment and interact online on the content of the microlecture through the Internet platform [23].

*2.2. Application of Microlecture in Ideological and Political Courses in Colleges.* The ideological and political courses are theoretical courses, and traditional teaching methods

are often difficult to achieve the expected results [24]. Due to the strong theoretical nature in the IaPC, students will find it difficult to learn [25]. In addition, a strong theory also means that the content is somewhat boring [26]. To solve the above issues, college teachers have tried a lot of means to reform teaching. For example, real stories are introduced into IaPC to increase the learning interest of students [27]. In addition, teachers are no longer limited to traditional offline teaching but began to explore teaching methods that combine IaPC with the Internet [28].

Microlecture meets the needs of college teachers for ideological and political curriculum reform in the mobile Internet era and is widely used in practical teaching [29]. First of all, microlecture can vividly explain ideological and political knowledge through video and audio methods, which solves the problem of boring content to a certain extent [30]. Secondly, microlecture can decompose a certain complex knowledge in a targeted manner to form continuous learning videos, which can reduce the learning difficulty of students [31]. Furthermore, with the increase of microlecture recording courses, a huge database of ideological and political knowledge can be formed, which facilitates students to search for specific knowledge [32]. Finally, microlecture makes up for the shortcomings of low participation of offline course students and insufficient teacher-student interaction [33].

From the above literature analysis, we can see that microlecture has played an important role in the innovation of IaPC in colleges. Scholars (e.g., Wang and Fan [30]; Cui [31]; Ran [33]) extensively discussed the advantages of microlecture in IaPC. However, the above-mentioned research is mainly based on conceptual and case studies and lacks empirical analysis. This paper is aimed at establishing a model of the influencing factors of microlecture on the teaching effect of IaPC in colleges and conducting empirical analysis on the data collected by the questionnaire and then identifying specific influencing factors.

## 3. Theoretical Framework and Research Hypothesis

*3.1. Theoretical Framework.* Based on the above analysis of the superiorities of microlecture and the discussion of the application of microlecture in IaPC, this study has determined the four influencing factors of microlecture on the teaching effect of IaPC in colleges, namely, single topic, easy to learn, convenient to share, and real-time interaction. Also, we constructed a model of the influencing factors of microlecture on IaPC in colleges, as shown in Figure 1. The theoretical model will guide the following research hypotheses, questionnaire design, and empirical analysis.

### 3.2. Research Hypothesis

*3.2.1. Single Topic.* Traditional IaPC usually adopt offline teaching methods. The time of a single course is relatively long, the knowledge points taught are more, and the topics involved are also diversified [24, 25]. The application of microlecture in IaPC adopts the opposite teaching concept.



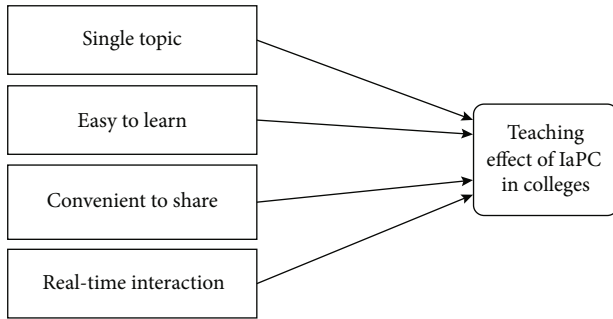


FIGURE 1: Theoretical model.

That is, each microlecture only teaches one topic. And when there is too much content in a certain topic, it tends to decompose the content of a single topic to form a set of topic content that can be continuously learned [4]. Its advantages are as follows: firstly, the subject is clear, which is convenient for students to search and review according to their learning needs. Students' learning provides confidence and convenience [31]. Based on the above discussion, we propose the following research hypotheses:

Ha: the single topic characteristics of microlecture are conducive to improving the teaching effect of IaPC in colleges.

**3.2.2. Easy to Learn.** IaPC are theoretically strong and involve many contents, and the teachers' teaching styles are boring, etc., which are often the main reasons why students are unwilling or not to actively participate in the ideological and political classroom [25, 26]. However, microlecture provides new ideas for solving the above-mentioned problems. The content decomposition concept of microlecture invisibly reduces the huge course pressure of students [31]. Introducing modern elements such as stories and news into the microlecture, as well as playing the microlecture repeatedly, can help students understand the theory very well [30]. In addition, the teacher's speech tone and speed can also be designed and optimized through computer software [34]. It is worth mentioning that microlecture does not move traditional offline classrooms to online. It can display ideological and political knowledge through animation and other forms, which can enhance learning autonomy and interest in learning [35]. Accordingly, we propose the following hypotheses:

Hb: the easy-to-learn characteristics of microlecture are conducive to enhancing the teaching effect of IaPC in colleges.

**3.2.3. Convenient to Share.** Traditional offline ideological and political teaching is a "one-off" course, while microlecture is recorded by means of video or audio [5] and can be disseminated on the Internet [3]. Particularly in the hot era of microvideo apps, microvideos of about 2 minutes consume less traffic and spread faster [36]. It is precisely because of the above characteristics that ideological and political knowledge points can also be integrated according to the current news, thus launching a microlecture that keeps pace with the times. The subject of the microlecture is single, and the content is small, and the teacher spends less time and

experience in making a microlecture [4]. After the microlecture is finished, it can be quickly uploaded to the Internet. And students can also search and forward on the Internet. Microlecture uses Internet platforms and mobile communication methods to make ideological and political teaching repeatable [37]. Accordingly, we propose the following research hypotheses:

Hc: the convenient-to-share characteristics of microlecture are conducive to improving the teaching effect of IaPC in colleges.

**3.2.4. Real-Time Interaction.** Teacher-student interaction is an important part of IaPC. The interactive part of traditional IaPC is in the "45-minute classroom," while the teacher-student interaction of the microlecture is based on the Internet and real-time [33]. After the teacher finishes making the ideological and political microlecture, they can upload it to the microvideo app (e.g., Douyin and Weishi) or share it with the class group. After watching the microlecture, students can like, forward, or comment online [38]. In addition, students can also share microlecture to WeChat Moments, Weibo, and other platforms. In this way, students can not only interact with teachers and classmates on a certain ideological and political knowledge. Because the user groups of different online platforms are different, students can also discuss and interact with groups of different identities (e.g., workers, natural scientists, and politicians) [39]. Obviously, the real-time interactive function of microlecture has expanded the scope of IaPC and increased the interaction between students and teachers, students and students, and even students and social groups. Therefore, we propose the following research hypotheses:

Hd: the real-time interaction characteristics of microlecture are conducive to enhancing the teaching effect of IaPC in colleges.

## 4. Research Methods and Results

**4.1. The Design of Questionnaire.** Based on literature analysis, model construction, and research hypotheses, we designed the research questionnaire. The questionnaire contains two parts, one of which is the basic information of the surveyed person (including gender, age, whether the respondent is a college student, whether the respondent has studied IaPC, and whether the respondent has any experience in learning through microlecture), and the other one part is the research variables (including single topic, easy to learn, convenient to share, real-time interaction, and teaching effect of IaPC in colleges). The basic information of the respondents is used as the control variable in this study. The observed variables of each latent variable were quantified using the 5-point Likert scale [40]. According to scholars' opinions and research conclusions, this paper developed the questionnaire scale (see Table 1).

**4.2. Data Collection.** Our data were collected through questionnaires. The objects of data collection are college students in our work unit, and IaPC are a compulsory course for them. The questionnaire survey started in June 2021 and

TABLE 1: Questionnaire scale.

Latent variables	Observed variables	References
Single topic (ST)	a1: I like the single topic of ideological and political knowledge learning method. a2: the single topic of the microlecture reduced my study pressure. a3: the single topic of the microlecture facilitates me to retrieve related ideological and political knowledge.	Zhao and Yang [4]; Cui [31]; Zhao and Zhan [32]
Easy to learn (EtL)	b1: the learning method of microlecture makes it easier for me to understand ideological and political knowledge. b2: I think the introduce microlecture into ideological and political teaching has made our study more interesting. b3: I think the teacher's lectures in microlecture are more vivid and interesting.	Wang and Fan [30]; Liu [34]; Lai and Pan [35]
Convenient to share (CtS)	c1: in my opinion, the fast-spreading speed of microlecture helps me to get in touch with more ideological and political knowledge. c2: I think it is convenient for me to share microlecture, which is conducive to my timely acquisition of relevant ideological and political knowledge. c3: in my opinion, the sharing of microlecture is convenient, which is conducive to my repeated learning of relevant ideological and political knowledge.	Zhou et al. [36]; Du [37]
Real-time interaction (RTI)	d1: I often make comments on the ideological and political content in the microlecture. d2: in my opinion, commenting on the ideological and political content in microlecture is conducive to the interaction between teachers and students. d3: in my opinion, microlecture provides a very good online platform for the interaction of ideological and political teaching.	Duan [38]; Lu [39]
Teaching effect of IaPC in colleges (TE)	h1: I think teachers can make full use of microlecture to improve the teaching effect. h2: I was more active in the microlecture. h3: I hope teachers can often teach IaPC in the way of microlecture.	Wen and Zhang [8]; Min [9]

TABLE 2: Descriptive statistical results of questionnaire survey.

Basic information	Items	Rate	Basic information	Items	Rate
Gender	Male	56%	College students	Yes	94%
	Female	44%		No	6%
Age	18-20	13%	Whether have studied the course of ideological and politics	Yes	100%
	21-23	58%		No	0%
	23-25	25%	Whether have the experience of learning through microlecture	Yes	95%
	>25	4%		No	5%

ended in July 2021. The questionnaire was distributed online. All respondents participated anonymously and were told at the beginning of the questionnaire that the questionnaire data were only used for scientific research. A total of 253 questionnaires were collected in this survey, among which 237 were valid, with an effective rate of 93.7%. The descriptive statistical results of this survey are shown in Table 2.

### 4.3. Data Analysis

**4.3.1. Reliability and Validity Test.** In this paper, SmartPLS software was used to verify the reliability and validity. The results in Table 3 show that Cronbach's Alpha (CA), Combined Reliability (CR), and Average Variance Extracted

TABLE 3: Results of reliability and validity tests.

Construct	CA	CR	AVE
ST	0.808	0.876	0.729
EtL	0.754	0.813	0.675
CtS	0.703	0.718	0.642
RTI	0.843	0.867	0.727
TE	0.735	0.768	0.695

(AVE) all reach corresponding thresholds [41], where the minimum value of CA is 0.703, the minimum value of CR is 0.718, and the minimum value of AVE is 0.642. The results in Figure 2 show that the external load values of each

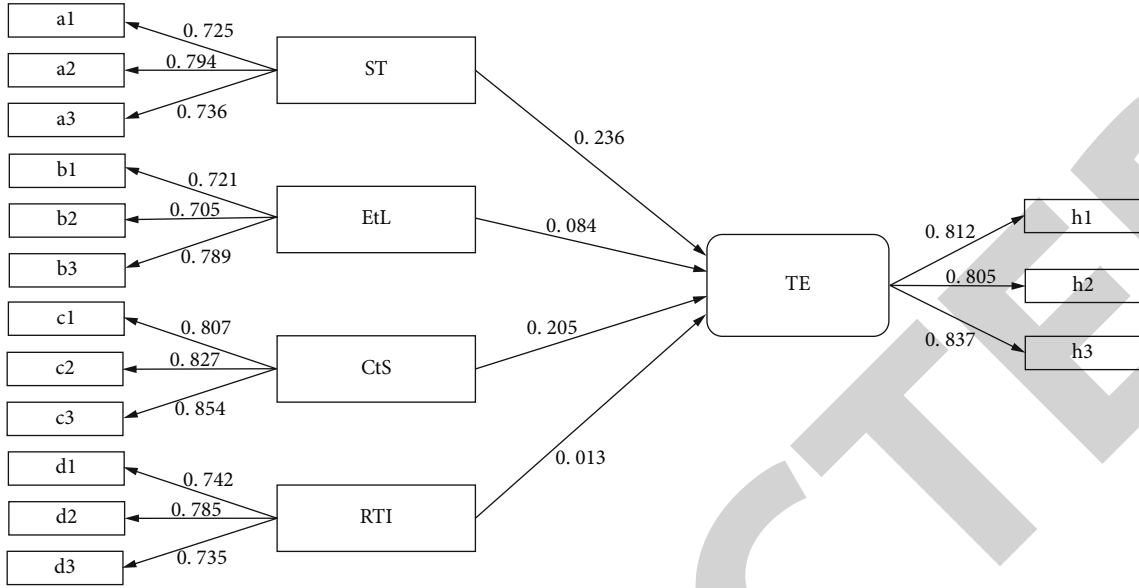


FIGURE 2: Results of the model.

TABLE 4: Average Variance Extracted (AVE) square root and factor correlation coefficient.

	H	Ha	Hb	Hc	Hd
TE	0.857				
ST	0.525	0.819			
EtL	0.407	0.627	0.762		
CtS	0.579	0.516	0.593	0.731	
RTI	0.431	0.348	0.375	0.426	0.658

Note: diagonal elements are the square root of Average Variance Extracted (AVE).

TABLE 5: Results of hypothesis testing.

Hypothesis	Path	Path coefficient	t value	P value	Hypothesis supported?
Ha	ST → TE	0.236	3.215	0.014	Y
Hb	EtL → TE	0.084	1.983	0.048	Y
Hc	CtS → TE	0.205	5.906	0.001	Y
Hd	RTI → TE	0.013	0.636	0.713	N

observed variable are all greater than 0.7, and the minimum value is 0.705. This indicates that the model we constructed has high reliability and validity [41]. The results in Table 4 show that the square root of AVE is greater than the correlation coefficients between it and other latent variables, which indicates that there is no multicollinearity between the latent variables in our study [42]. The built model has an  $R^2$  value of 0.47, indicating that it has excellent explanatory power. In addition, the model's fitting degree has reached the required level.

4.3.2. *Hypothesis Testing.* We performed bootstrapping operation in SmartPLS software to verify the research hypothesis. The results showed (see Table 5) that Ha, Hb, and Hc have

passed the test ( $P \leq 0.05$ ) [42], while Hd did not pass the test ( $P > 0.5$ ) [42]. The results show that in microlecture teaching, single topic, easy to learn, and sharing convenience are the main factors to enhance the teaching effect of IaPC in universities, while real-time interaction had no significant effect. Therefore, we find that factors such as single topic (Ha), easy to learn (Hb), and sharing convenience (Hc) support the conclusions and opinions of previous scholars. However, our research shows that real-time interaction (Hd) is contrary to the views of Duan [38] and Lu [39].

## 5. Conclusions

The offline teaching mode of IaPC is difficult to stimulate students' enthusiasm for learning because of its over content and strong theory. The emergence of microlecture offers a novel way to figure out the above problems. Microlecture has the characteristics of single topic, easy to learn, convenient sharing, and real-time interaction. Whether these characteristics are conducive to enhancing the effect of ideological and political teaching in colleges is the central point of this paper. Therefore, the focus of this study is on identifying factors that influence the teaching effect of microlecture in college ideology and politics courses. In particular, we developed a microlecture influence factor model for ideological and political education in colleges. We confirm the four features of microlecture by a questionnaire survey and empirical research. The results show that in microlecture teaching, single topic, easy to learn, and sharing convenience are the main factors to enhance the teaching effect of IaPC in colleges, while real-time interaction had no significant effect.

Our research has contributed to the literature on microlecture in ideological and political teaching by identifying the possible factors influencing ideological and political teaching in colleges, as well as confirming that microlecture can help improve the teaching effect of ideological and political in colleges. Based on the research results, we propose the following

recommendations: (1) in ideological and political teaching, recorded microlecture should concentrate on a single topic as far as possible, and the time of microlecture should be controlled to be shorter; (2) since students believe that microlecture makes their study easier and share convenience, universities should consider promoting this mode in IaPC.

However, there are some limitations in this study. First, from the characteristics of microlecture, we have discussed their influence on the teaching effect of IaPC in colleges. But factors such as the level of use of smart mobile devices and the prevalence of the Internet could also influence the results. Therefore, future study can be conducted based on more diverse perspectives and consider more factors. Second, we have explored the influencing factors of microlecture in the teaching of IaPC, but the influence of different courses needs to be further explored. Third, the samples in this study are mainly from our work units, and it remains to be further proved whether the research results are different among other universities. Fourth, because our data is collected online, it is difficult to trace the accuracy of the information provided by respondents. Hence, in the future, we may be able to gather data offline or include a test to verify the accuracy of the information in the questionnaire. Finally, the sample size of effective questionnaires collected in this paper is limited, so the sample size can be appropriately increased in the future, so as to obtain more general results.

## Data Availability

The data used to support the findings of this study are included within the article.

## Conflicts of Interest

The author does not have any possible conflicts of interest.

## References

- [1] C. Herreid, N. Schiller, K. Herreid, and C. Wright, "Case study: a chat with the survey monkey: case studies and the flipped classroom," *Journal of College Science Teaching*, vol. 44, no. 1, pp. 75–80, 2014.
- [2] C. F. Herreid and N. A. Schiller, "Case study: case studies and the flipped classroom," *Journal of College Science Teaching*, vol. 42, no. 5, pp. 62–67, 2013.
- [3] D. Sweet, "Microlectures in a flipped classroom: application, creation and resources," *Mid-Western Educational Researcher*, vol. 26, pp. 52–59, 2014.
- [4] H. Zhao and C. Yang, "Application of microlectures based on WeChat in rehabilitation nursing teaching," in *2019 10th International Conference on Information Technology in Medicine and Education (ITME)*, Qingdao, China, 2019.
- [5] Educause, "7 things you should know about microlectures," <http://www.educause.edu/library/resources/7-things-you-should-know-about-microlectures>.
- [6] P. Zheng, X. Wang, and J. Li, "Exploration and practice of curriculum ideological and political construction reform — take "information security" course as an example," *ASP Transactions on Computers*, vol. 1, no. 1, pp. 1–5, 2021.
- [7] S. Liu and S. Wu, "Limitations and options of field work for courses of ideological and political education," *Teaching and Research*, vol. 4, pp. 87–90, 2008.
- [8] C. Wen and J. Zhang, "Design of a microlecture mobile learning system based on smartphone and web platforms," *IEEE Transactions on Education*, vol. 58, no. 3, pp. 203–207, 2015.
- [9] M. Weina, *Application of "Micro-Lectures" of the Courses about Ideological and Political Theory in Science and Engineering Universities from Perspective of Big Data and Considerations*, Advances in Social Science, Education and Humanities Research, 2018.
- [10] J. Zeng, "Comprehensive application of microlecture technology in college ideological and political education," *Journal of Physics: Conference Series*, vol. 1578, no. 1, article 012068, 2020.
- [11] W. Wu, "Discussion on the integration of innovation and entrepreneurship education and ideological and political education in colleges and universities from the perspective of "Internet plus"," *Journal of Frontiers in Educational Research*, vol. 1, no. 4, pp. 137–142, 2021.
- [12] J. McCormick, K. Vincs, S. Nahavandi, D. Creighton, and S. Hutchison, "Teaching a digital performing agent: artificial neural network and hidden markov model for recognising and performing dance movement," in *Proceedings of the 2014 International Workshop on Movement and Computing - MOCO '14*, pp. 70–75, Paris, France, 2014.
- [13] H. Li, "A teaching quality evaluation model based on a wavelet neural network improved by particle swarm optimization," *Cybernetics & Information Technologies*, vol. 14, no. 3, pp. 110–120, 2014.
- [14] S. Thanikodi, D. K. Singaravelu, C. Devarajan, V. Venkatraman, and V. Rathinavelu, "Teaching learning optimization and neural network for the effective prediction of heat transfer rates in tube heat exchangers," *Thermal Science*, vol. 24, no. 1, Part B, pp. 575–581, 2020.
- [15] Y. A. Ahmed and K. Hasegawa, "Automatic ship berthing using artificial neural network trained by consistent teaching data using nonlinear programming method," *Engineering Applications of Artificial Intelligence*, vol. 26, no. 10, pp. 2287–2304, 2013.
- [16] J. Zhang, Y. Liu, H. Liu, and J. Wang, "Learning local–global multiple correlation filters for robust visual tracking with Kalman filter redetection," *Sensors*, vol. 21, no. 4, p. 1129, 2021.
- [17] J. Zhang, J. Sun, J. Wang, and X. G. Yue, "Visual object tracking based on residual network and cascaded correlation filters," *Journal of ambient intelligence and humanized computing*, vol. 12, pp. 8427–8440, 2021.
- [18] M. Xiangzeng, L. Ruimei, and W. Guangxin, "The theory and practice of the design and construction of microlectures," *Journal of Distance Education*, vol. 6, pp. 24–32, 2014.
- [19] L. Huang, G. Xie, J. Blenkinsopp, R. Huang, and H. Bin, "Crowdsourcing for sustainable urban logistics: exploring the factors influencing crowd workers' participative behavior," *Sustainability*, vol. 12, no. 8, p. 3091, 2020.
- [20] Z. Wang, Y. Luo, and Y. Qu, "Application of micro-lecture for engineering mechanics experimental teaching," *International Journal of Innovation and Research in Educational Sciences*, vol. 2, no. 4, pp. 130–132, 2017.
- [21] W. Zhang, "Design a civil engineering micro-lecture platform based on the ARCS model perspective," *International Journal of Emerging Technologies in Learning*, vol. 12, no. 1, 2017.

## Research Article

# Construction and Simulation of a Multiattribute Training Data Mining Model for Basketball Players Based on Big Data

Yunbin Li<sup>1</sup>, Jinyan Ge<sup>1</sup> and Wei Hao<sup>2</sup>

<sup>1</sup>*Xi'an Jiaotong University Physical Education Centre, Xi'an 710049, China*

<sup>2</sup>*P.E. Department of Northwestern Polytechnic University, Xi'an 710072, China*

Correspondence should be addressed to Yunbin Li; [11137332@stu.wxjic.edu.cn](mailto:11137332@stu.wxjic.edu.cn)

Received 4 August 2021; Revised 30 August 2021; Accepted 1 September 2021; Published 30 September 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Yunbin Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper provides an in-depth analysis and research on the construction and simulation of a big data model for multiattribute training of basketball players. To get a more accurate and three-dimensional information, the training can use a multitraining target robot, i.e., to detect feedback on multiple indicators at the same time and correct the player's errors in time; the other is an auxiliary robot, which can actively correct technical movements and train the player to form muscle memory, compared with general training. The analysis results show that by either constructing a human model or designing an active assistive robot, the player's technical movements can be regulated accordingly, protecting the player's body laterally and improving the player's ability. An assisted training system with an accurate model of physiological indicators is constructed based on the data of the player throughout the season. The Warriors, who have applied this system, not only have the best record in recent years but also have the lowest injury rate in the league, indicating that this method has indeed reduced the injury rate of players.

## 1. Introduction

Shooting is a general term for the various methods of action used by a player with the ball to put the ball into the basket from the top of the rim using various correct techniques, and the cumulative number of shooting hits determines the winner of the game. Therefore, shooting becomes the focus of competition between offense and defense and is the core technology of basketball. With the development of basketball, the athletes' morphology and functional quality improve, prompting the continuous development of shooting technology, from low to high shooting parts, from slow to fast shooting speed [1]. This evolution and development of the shooting technique are marking the leap from one stage to another in basketball. An athlete's daily training includes physical training, shooting, rebounding, defense, and fast break, of which shooting training is the longest process of basketball skills. Shooting training, in turn, contains free throw shooting, jump shooting, three-point shooting, and other training [2]. The free throw drill, as the basic shooting drill, is the beginning of that skill training. Once

any irregular technical movements of players in this stage are developed, it will directly affect their inconsistent shooting percentage during the game and prevent them from further improving their shooting percentage [3]. Therefore, in the free throw training stage, it is especially important to develop the shooting technique and rhythm habit of free throw shooting. There are an increasing number of shooting methods, and the hit rate continues to increase. This evolution and development of shooting technology is marking the leap of basketball from one stage to another. The daily training of athletes includes physical training, shooting, rebounding, defense, and fast break. Among them, shooting training is the longest running-in process in basketball skills. At the same time, in the game, due to the high intensity and high confrontation, a foul that brings free throws is the most simple and effective means of scoring, often in the late stages of the game, affecting the trend of the whole game and even the game's victory or defeat.

The fundamentals of competitive basketball are in developing and creating the style of basketball and in the process of the continuous development of the style to enhance their

space of understanding. As the times evolve, the style presented by the game of basketball will become the brand of basketball of that era. Style is the research crystallization and core of the winning elements of the game [4]. The gradual improvement in the understanding of the laws of basketball, the understanding of the game-winning factors from shallow to deep, and the emergence of ways to defeat the opponent have driven the display of various styles. Sports style is the core content of basketball culture, more to show a spiritual culture connotation. It is the driving force that keeps basketball alive. Quantitative research usually uses numbers to measure the object of the study and find the connection between variables with the help of mathematical and statistical means. Quantitative research can represent some nonspecific, vague factors with specific data to conclude the overall trend or regularity. In turn, the patterns can be generalized [5]. Basketball is a competitive sport in which points are scored by putting the ball in the basket. To win the game, you must hit as many baskets as possible in as much an effective time as possible. Therefore, shooting is the most important training subject in daily training, and it is also the most skillful and longest training program in terms of practice accumulation. A good shot requires not only to put the ball in but also to be able to put the ball in the basket again in the face of defensive interference or fast movement. This requires the athlete to make full use of body coordination, stretch the body reasonably, and finally throw the ball in the most suitable state to throw the ball. Therefore, there are many research results on the analysis of shooting techniques.

For sports, young people are the future and are the backbone of accelerating the construction of a strong sports nation. The new physical education ecosystem is built based on new-generation information technology such as big data and artificial intelligence. Its core is to help students personalize their physical education learning and training, to help teachers provide scientific guidance on the technical and tactical aspects and special qualities of sports, to assist administrators in efficient and orderly school governance, and to achieve intelligent and coordinated communication among schools, families, and society. The main teaching objects facing school sports are primary and secondary school students, the future force of AI development. By imparting intelligent knowledge and using intelligent products in daily sports teaching, information literacy is enhanced to achieve the purpose of improving the AI science education system and providing talent reserves for the development of AI in China. The current application of intelligent technology in the field of sports, especially basketball, is still in the exploration stage, the systems and technologies are not yet mature, the purpose and scope of application are not yet clear, so it becomes especially important to find the combination of intelligent technology and the field of basketball and to develop and optimize the method of the intelligent application system for sports basketball. This paper focuses on the theme of the application of artificial intelligence in the field of basketball, using literature research, video analysis, comparative research, mathematical statistics, and other research methods to explore and analyze the implementation of arti-

ficial intelligence in basketball big data, basketball robotics, and basketball teaching and training methods, to provide a theoretical basis for promoting the application of artificial intelligence in the field of basketball, and also to provide a broader range of artificial intelligence in the field of sports. It also provides a theoretical reference for the wider application of AI in sports.

## 2. Current Status of Research

We humans call ourselves *Homo sapiens* because we have recognized that intelligence is an extremely important trait for us. We can perceive, understand, predict, and manipulate a world that is far larger and more complex than ourselves by relying on our brains alone [6]. This is the obvious and powerful role of intelligence. But for the field of artificial intelligence, what is needed is not just to try to understand intelligence but to try to create and design intelligence based on our understanding of intelligence [7]. Today, the field of artificial intelligence as we know it covers a wide variety of subfields and scopes, ranging from the treatment of persistent diseases to chess, to the proof of a mathematical theorem, to intelligent driving, and to our usual learning and perception [8]. Nowadays, the novel general artificial intelligence is relatively high-end in the field of intelligence, because not only is general artificial intelligence limited to a certain aspect of intelligence, but also nowadays, it is aimed at creating a system that is almost as intelligent as a human being; in other words, it is aimed at achieving full intelligence [9]. To win the game, you must hit the basket as many times as possible within the effective time. Therefore, shooting training is the most important training subject in daily training. It is also the training project with the strongest skills and the longest process of accumulation of practice. The cotermin analysis is a research method that indicates the relationship between topics by the number of common occurrences of condensed words that represent a piece of literature, and if the number of common occurrences is higher, the closer the relationship between the two topics. The cluster analysis method is a mathematical operation of the distance dimension on the relationship between objects in the complex network formed by the research objects, and the objects that are closer together are clustered into a major group, which is more like the objects within the group and more different from other taxa [10].

Along with the improvement of computer performance and image computing power, it is becoming increasingly common to apply techniques related to video images to sports analysis scenarios, such as when [11] used multiple cameras to cover the whole field of a soccer match, seeking to capture the on-field movements and trajectories of each player [11]. A tracking algorithm is proposed to obtain the trajectories of soccer players during a match. The algorithm can identify targets well by building individual models based on features and athletes' motion trends and can help coaches statistically analyze the tactical runs and movements of athletes during the game [12]. In a study by Li et al., a human body fitting model based on morphological techniques in a simple context was proposed to locate the four key joints

of the body and extract the angular information of the body and the angle of water entry to achieve the purpose of assisting divers in correcting their aerial movements and water entry posture in their daily training [13]. The method proposes an algorithm that automatically acquires lead shot exit frames [14]. The algorithm performs lead shot frame prediction, filters out images that do not contain shot frames in the image frame sequence, and then obtains the position of the lead ball by Hough transformation, further verifies and filters the shot frames, and eliminates the pseudo-shot frames before the shot frames. Goes et al. used a monocular camera to extract the 3D motion pose of the human body and applied it to identify and detect golfers' swings [15]. By comparing the detected posture information of the human skeleton with that of a professional athlete, appropriate guidance is given.

In general, there are few research articles on the application of knowledge graphs in sports science research, but they have shown rapid growth in recent years and have been applied to many fields in sports, achieving rich research results and demonstrating their applicability and effectiveness in sports science research. Many of the articles have a high gold content and are published in core journals or become research directions for experts and doctoral theses. This is both a recognition of the application of scientific knowledge mapping to sports research and a side indication that this research method still has great potential. They hope to realize a new way of interaction between gestures and eye-following movements.

### **3. Basketball Players' Multiattribute Training Big Data Mining Model Construction and Simulation Analysis**

*3.1. Big Data Mining Design for Multiattribute Training of Basketball Players.* The process of coming up with hidden, potentially valuable, and informative information from a massive database is called data mining. We can define data mining from different standpoints, and despite the differences in these expressions, there is no contradiction in its essence. In defining data mining, the authors focus on the technical perspective and the business perspective [16]. It provides a theoretical basis for promoting the application of artificial intelligence in the field of basketball sports and provides a theoretical reference for the wider application of artificial intelligence in the field of sports. From the technical point of view, data mining is the process of extracting and collecting hidden, but valuable, information and knowledge from a large amount of data with interference and randomness with the help of computer technology and other means. It is believed that knowledge takes data as an important source. The raw data existing in relational databases are structured, while the raw data existing in text, graphics, and images are semistructured. We can acquire knowledge by mathematical and nonmathematical means, and we can also acquire knowledge by deductive and inductive means. The knowledge we obtain through data mining can be used both for optimization and management of information and

at the same time for supporting decision-making. In summary, data mining is a multidisciplinary technology that brings together researchers from different fields. In the future, data mining will be the focus and hot spot for technology research and development.

Data mining makes use of statistics and computer technology to find patterns that fit market dynamics and customer psychology, and currently, data mining technology has been fully developed to automate mining techniques, while combining data mining and business data warehousing to show the results of mining to the decision-makers of the company. We can acquire knowledge through mathematical and nonmathematical means, and we can also acquire knowledge through deduction and induction. The knowledge we obtain through data mining can be used to optimize and manage information and, at the same time, can also be used to support decision-making. In the process of applying data mining techniques, it is not only necessary to make use of scientific algorithms and build data mining models. At the same time, it is necessary to ensure that data mining technology does not conflict with the complex information technology application environment. On the other hand, data mining technology requires the participation of data mining analysts [17]. In contrast, data mining technology is not intuitive and cannot judge the meaning of the mining model in real life, so the participation of data mining analysts should be valued. Since the amount of data to be acquired exceeds 10 GB, the load balancing capability and stability of the system are highly required, and the crawler module involved needs to be implemented using a distributed architecture, and the data is stored uniformly. In addition, the problem of rescheduling tasks due to the short-term failure of crawler nodes and the problem of how to replace crawler nodes with long-term failure are also issues that the system needs to focus on.

High-efficiency requirements, due to a large amount of data and heavy crawling task, will require the system to take into account the efficiency problem under the premise of a stable operation, which is the need to use the agent node technology to solve and complete continuous data acquisition, but because the agent node technology will be with the help of a third party, while the service has limited thread work capacity, so the technology used by the agent node will also need to consider the load balancing problem; in addition, the overall data calibration project will also consume larger problems and put higher requirements on the accuracy of system crawling, as shown in Figure 1.

A foul against sportsmanship by a player is awarded to the fouled pair two free throws and one serve; if the fouled player does not make the shot, but the shot is missed, two or three free throws and one server are awarded. Technical foul: a technical foul by a player awards the opposing team two free throws and a right to serve; the same result is awarded for technical fouls or unsportsmanlike fouls by coaches, substitutes, or accompanying personnel. A camera is set up on the extension of the free throw line to record the player's free throw shooting action. For the convenience of the analysis and the stability of the free throw technique, 20 free throws were captured at a time as a basic video

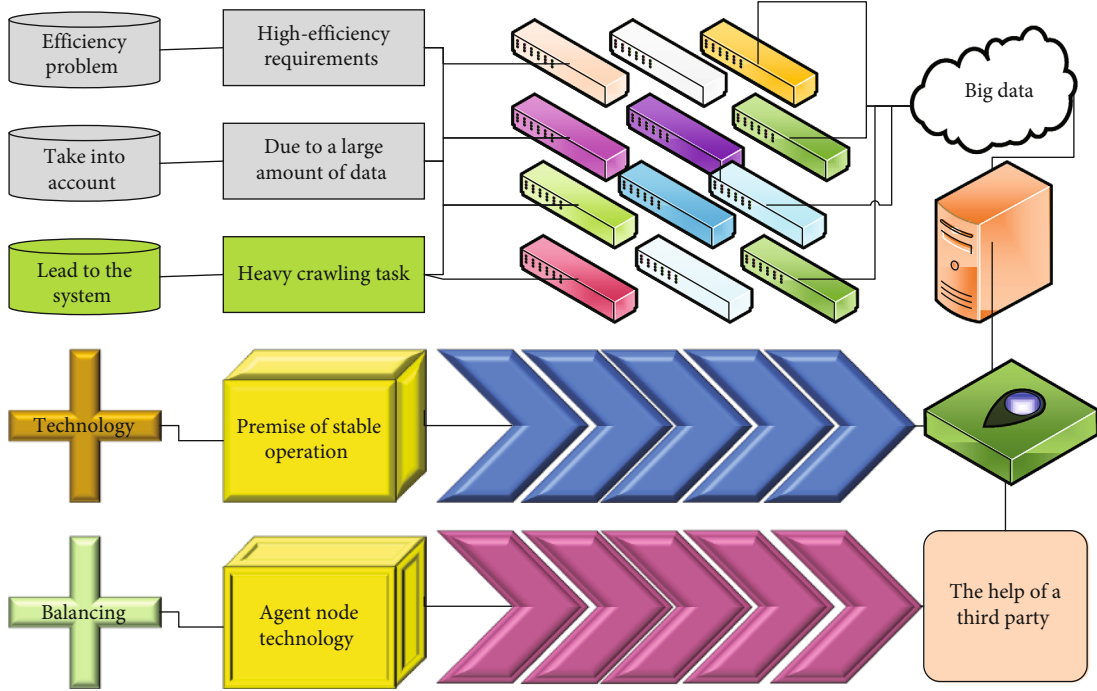


FIGURE 1: Multiattribute basketball training big data design framework.

analysis unit in the thesis. The background can be approximated as static because the relative spatial displacement of the athlete is small during the execution of the free throws at the free throw line position.

The basketball's shot information, which determines the flight trajectory, corresponds to a shot angle and a shot speed at the time of the shot and a corresponding entry angle and entry speed at the time of the basket. The angle of entry is the angle between the tangent line and the plane of the basket at the point of the trajectory now at entry when the center of the ball passes through the plane of the basket. Data mining technology requires the participation of data mining analysts. In contrast, data mining technology is not intuitive and cannot judge the meaning of mining models in real life. Therefore, the participation of data mining analysts should be emphasized. When the center of gravity of the ball passes through the center of the circle of the basket, it is a "hollow entry." Theoretically, the ideal angle of entry is  $90^\circ$ , which has the highest hitting rate and allows the largest margin of error. However, as the angle of entry decreases, the permissible error range becomes smaller. In the following, we define  $E$  to represent the allowable error in basketball entry,  $R$  is the radius of the basketball frame (22.5 cm), and  $r$  is the radius of the basketball (12-13.5 cm), then we have

$$E = R \cos \alpha + r. \quad (1)$$

Then, the relationship between the angle of entry  $\alpha$  and  $R$  and  $r$  is as follows:

$$\cos \alpha = \frac{r}{R}. \quad (2)$$

When taking the value of  $r$  as the middle value, we can get  $\alpha = 32.39^\circ$ . At this point, if the angle of entry is reduced again, the basketball will have to touch the front of the basketball frame and not directly into the frame. The corresponding allowable error is 0, so the angle of entry has a range:  $32.39^\circ - 90^\circ$ . A good angle of entry is the key to the hitting rate, but a good angle of entry is achieved by the control of the player of the angle of the ball. The following is an analysis of the basketball's flight trajectory. The ball's flight trajectory can be considered as a parabola by ignoring the forces other than gravity during the ball's flight and considering that the ball is only subjected to the action of gravity in the air. Based on physical knowledge, an equation between the factors can be derived as follows:

$$V_0^2 = \frac{gL}{2 \sin^2 \theta (\cot \alpha + h/L)}. \quad (3)$$

The speed of the shot is determined by four parameters: the shot height, the shooting distance, the angle of the shot, and the angle of the basket, and for a more detailed analysis, the paper makes the following analysis by setting the relevant parameters. Setting a certain shooting distance, the ball is subjected to the action of gravity only ignoring all other resistances, and the components of the ball's exit velocity in the horizontal and vertical directions are as follows.

$$\begin{cases} V_x = V_x \sin \theta, \\ V_y = V_x \cos \theta. \end{cases} \quad (4)$$

Obviously, the greater the angle of entry  $\alpha$ , the better, but an increase in  $\alpha$  requires the ball to gain greater speed and



angle of exit at the time of the shot, undoubtedly increasing the difficulty of the shot. Good free throw shooting technique training is to get the ball in with the least effort, which is especially important in the late game of basketball when physical exertion is too high. The following theoretical analysis is used to derive the possibility and existence of minimum speed and angle of the shot. In the free throw process, the shooting distance and the height of the shot are determined. The Vom required to hit the center of the basket with an  $\theta$  angle of the shot is minimal; this angle is called  $\theta$  which is the minimum velocity angle. Correspondingly, the allowable error at this time is 0, so the basket angle has a range of  $32.39^\circ$ - $90^\circ$ . The muscles gain memory through a lot of repetitive training and can always be adjusted with a better power posture and shot angle to obtain an improvement in free throw shooting rate to obtain a guarantee of free throw shooting rate when it matters. The theoretical optimum provides the direction of improvement for the athlete's daily training, as shown in Figure 2.

We analyze the players' playing time, physical exertion, and change of status and then make reasonable substitutions to improve the team's defensive or offensive capabilities [18]. As we all know, as the game match goes on, different players' statuses and physical exertions are different. By establishing a video database, the physical condition of both players can be monitored in real-time to improve the overall defensive and offensive capabilities. In training, chips and sensors are used to accurately record players' strength when shooting, running speed, and holding time, etc. These data are compiled and analyzed by analysts and coaching staff, from which the players' technical characteristics and physical conditions are judged, and on this basis, training plans are developed to meet the players' needs.

**3.2. Mining Model Construction and Simulation Experiment Design.** Holdout validation is not a kind of cross-validation in the strict sense. It is based on the principle of randomly dividing the dataset into two parts, one for training the model and the remaining part for validating the model, without forming a cross-validation process in the process. Its advantage is that it can avoid the practice of training and testing on the same dataset at the same time, thus reducing the phenomenon of model overfitting caused by the reuse of training and testing data, but because it groups the original data by a random method, the final model validation effect will depend excessively on the method of grouping data, which is not conducive to widespread application in actual prediction [19]. Take each subset of data as a validation set, and the remaining  $K - 1$  subsets of data will be used as the training set. In this way,  $K$  models will be obtained, and the average of the classification accuracy of the final validation set of these  $K$  models will be used as this  $K$ . The performance index of the classifier is under CV, which is a  $k$ -fold cross-validation basic validation. The method is as follows: the original data is divided into  $K$  groups, each subset of data is used as a validation set, and the remaining  $K - 1$  subsets are used as training sets so that  $K$  models are obtained, and the average of the classification accuracy of the final validation set of these  $K$  models is used

as the performance index of the classifier under  $K$ -CV.  $K$ -CV is the most commonly used cross-validation method, which can effectively avoid the occurrence of overfitting and underfitting, and the results obtained are more convincing if  $K$  is larger, and  $K$ -CV also greatly improves the computational cost from the performance perspective, as shown in Figure 3.

Accuracy is the most basic measure of the evaluation of the binary classification model and is calculated by the formula shown in (5), which denotes the number of TP and TN divided by the number of all predicted values, and TP, TN, FP, and FN are the number of times the predicted values of the model fall into these categories.

$$\text{Acc} = \frac{\text{TP} - \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \quad (5)$$

The error rate is the proportion of incorrect classifications and is calculated as shown in (6). The error rate can also be obtained by subtracting the accuracy rate from 1.

$$\text{Error} = 1 - \text{Acc}. \quad (6)$$

A useful classifier often must make a good trade-off between too conservative and too aggressive decisions. This trade-off can be achieved by two metrics: sensitivity and specificity. The sensitivity of the model is used to measure the proportion of positive samples that are correctly classified, as shown in

$$\text{Correctly} = \frac{\text{TP}}{\text{TP} - \text{FN}}. \quad (7)$$

The specificity of the model is used to measure the proportion of negative samples that are correctly classified, as shown in

$$\text{Negative} = \frac{\text{TN}}{\text{TN} - \text{FP}}. \quad (8)$$

The range of values for both sensitivity and specificity is 0 to 1, with values closer to 1 being more satisfactory. It is important to find an appropriate balance between the two, but this balance is often determined on a case-by-case basis. Two evaluation metrics, prediction accuracy, and recall accuracy were initially used in the field of information retrieval to describe the relevance of the model results or whether the predictive power of the model would be diminished by meaningless noise. Predictive accuracy is defined as the proportion of true positives among all predicted positive cases, and an accurate model will predict positives only when the category is very much like a positive, in which case the prediction is very reliable. An accurate model will only predict positive when the category is very similar to positive. In this case, the prediction is very reliable. On the other hand, retrospective accuracy is a measure of the completeness of results, which is defined as the ratio of true positives to the total number of negatives. Although this is the same formula as sensitivity, its meaning is different, and a model with high retrospective accuracy can capture many positive

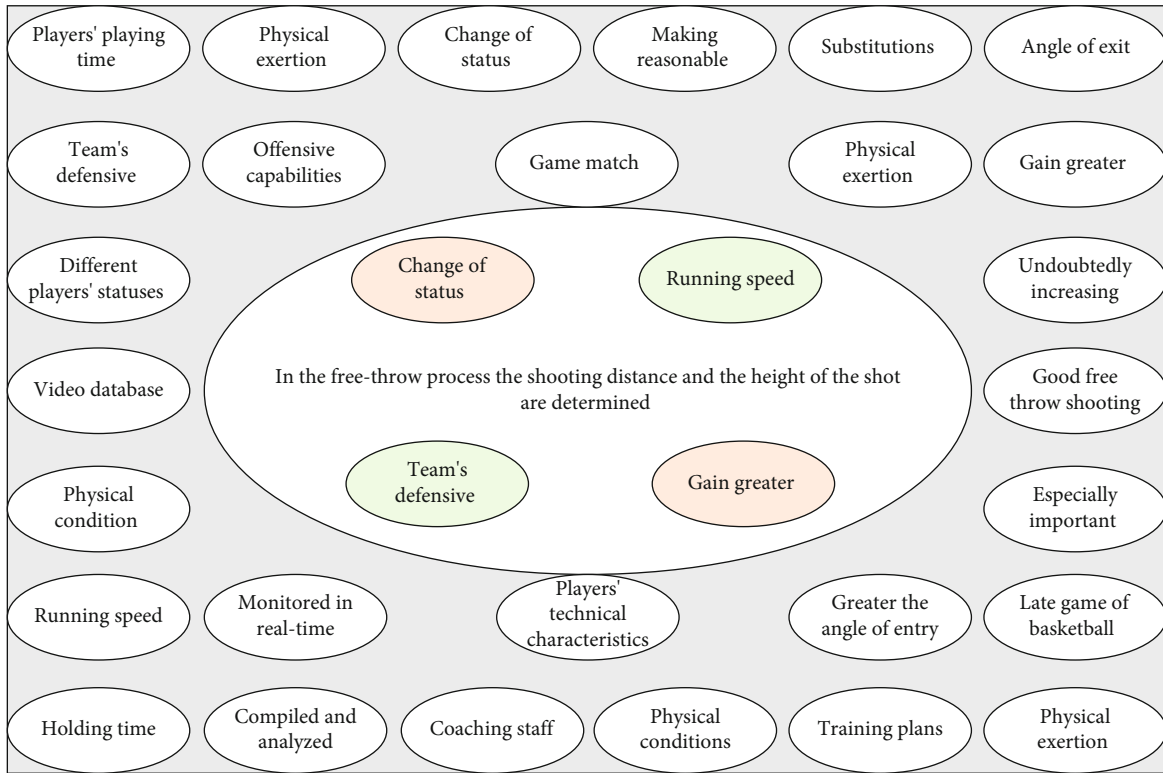


FIGURE 2: Multiattribute training feature design.

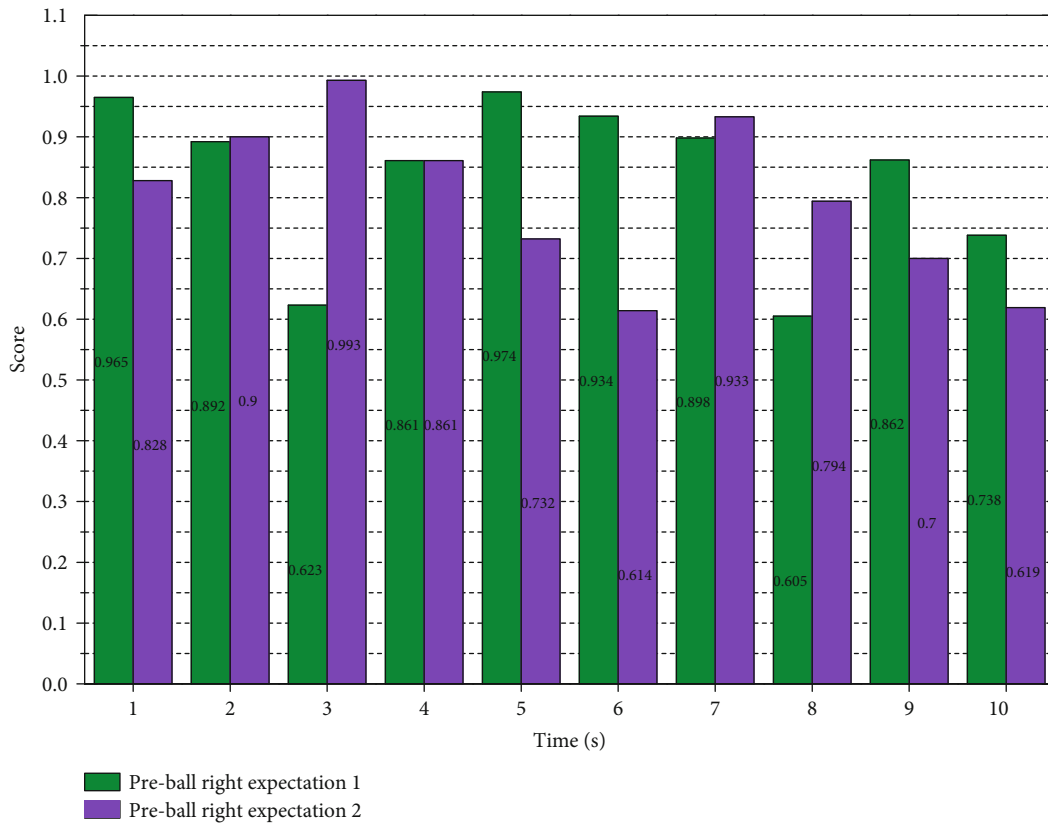


FIGURE 3: Preball possession score and time statistics.

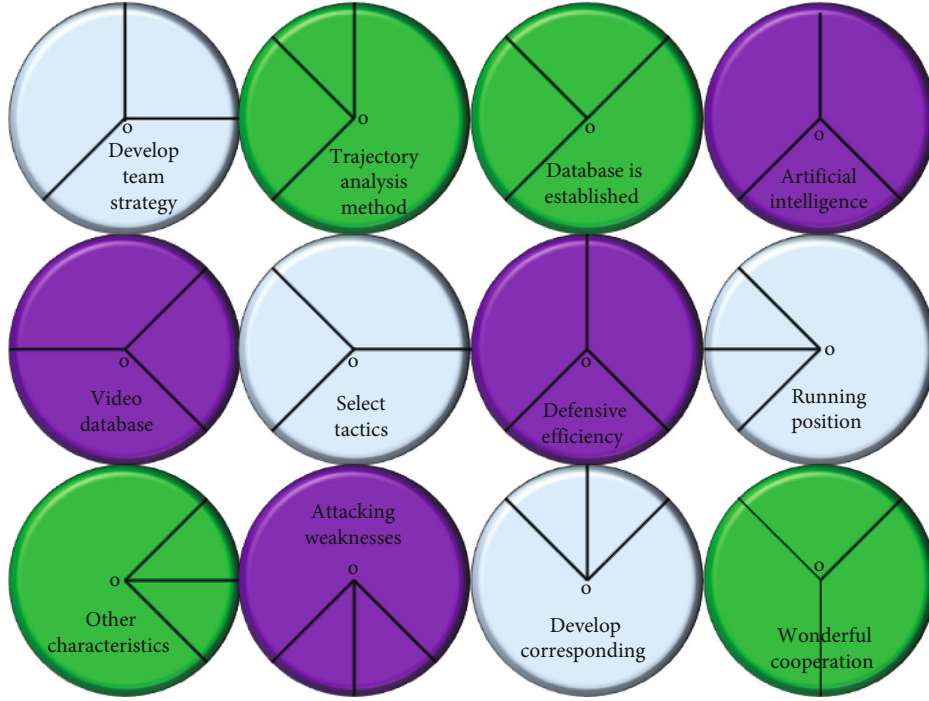


FIGURE 4: Schematic analysis of wrist paddle action.

samples, which means it has a wide range of applications. The specific formulas for prediction accuracy and retrospective accuracy are shown in

$$PAcc = \frac{TP}{TP + FN}, \quad (9)$$

$$GAcc = \frac{TN}{TN + FN}. \quad (10)$$

A model performance measure that combines predictive accuracy and retrospective accuracy into a single value is the  $F$ -metric, which uses a harmonic mean to integrate predictive accuracy and retrospective accuracy. The summed mean is an average value used to describe the rate of change, and because both predictive and retrospective accuracies have a range of 0 to 1, the summed mean is used instead of the more common arithmetic mean. The specific formula for calculating the  $F$ -metric is shown in equation (11).

$$F = \frac{2TN}{TN - FN}, \quad (11)$$

$$\theta_m = 45^\circ - \frac{1}{2} \cot^{-1} \left( \frac{h}{L} \right), \quad (12)$$

$$V_{0m} = \left[ g \left( h - \sqrt{h^2 - L^2} \right) \right]^2, \quad (13)$$

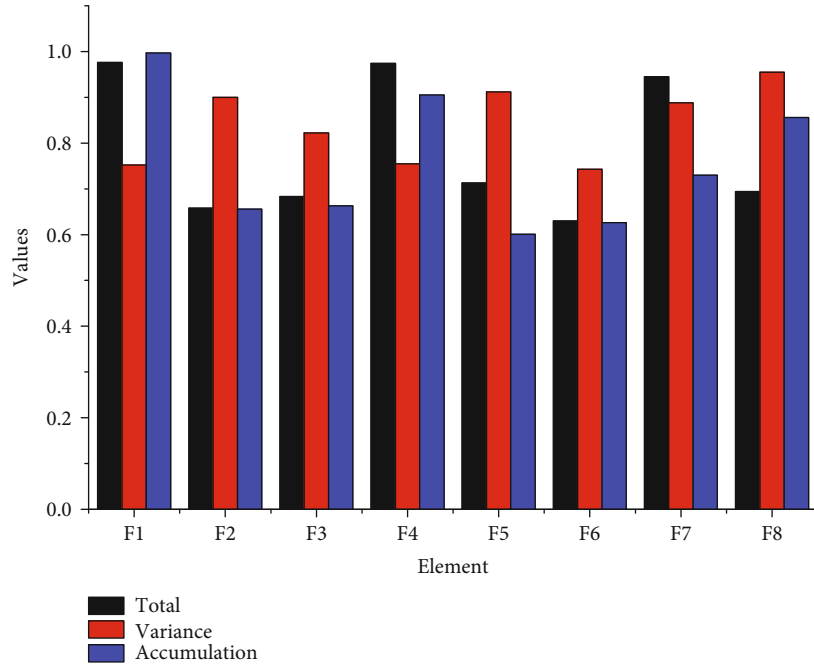
$$V_{in-y} = \frac{Lg}{V_0 \sin \theta} + V_0 \cos \theta, \quad (14)$$

$$\cot \alpha = \frac{Lg}{V_0 \sin^2 \theta} + V_0 \cos^2 \theta. \quad (15)$$

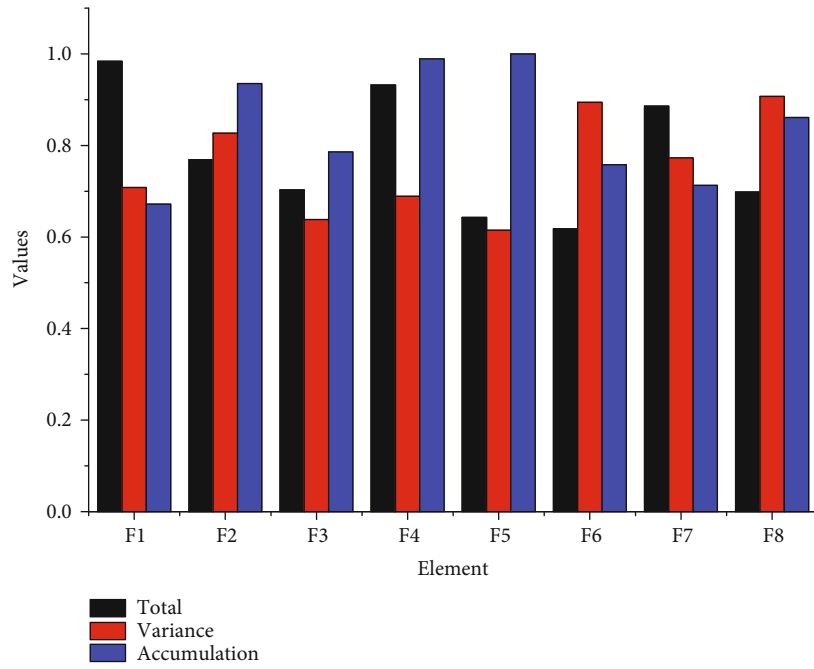
In summary, the accuracy rate is used as the first evaluation index to judge the classification model of this paper, and the  $F$ -measure will be used as the second evaluation index because the classification model of this paper satisfies the premise that the prediction accuracy and the retrospective accuracy have the same weight, and the other indexes are used as auxiliary indexes.

*Develop team strategy using trajectory analysis method:* after the database is established, the trajectory of the target is analyzed by artificial intelligence of big data to extract multiple movement trajectories in the video database of the opposing team and select tactics with higher offensive and defensive efficiencies. Players' running position, speed, characteristics, and game habits, footwork, blocking, and other characteristics to find out the opponent's attacking weaknesses are captured, and corresponding defensive strategies are developed [20]. *Player joint analysis method to develop team tactics:* basketball is a high-level team sport. On the court, we often see the wonderful cooperation between players. The cooperation between different players will produce different offensive efficiencies. In the database, the players who often cooperate on the court are identified, and the offensive or defensive efficiency of multiple players' partnership is analyzed from different time conditions (such as playing time and physical exertion), and the best lineup of the team is identified to enrich the team tactics and establish the best team tactical system, as shown in Figure 4.

The main purpose is to restore the core of the whole technical movement of the free throw basketball by marking the joint points of the arm and analyzing the trajectory of the marked feature points. The research objectives in the thesis can be summarized into two parts. Let the athletes maintain

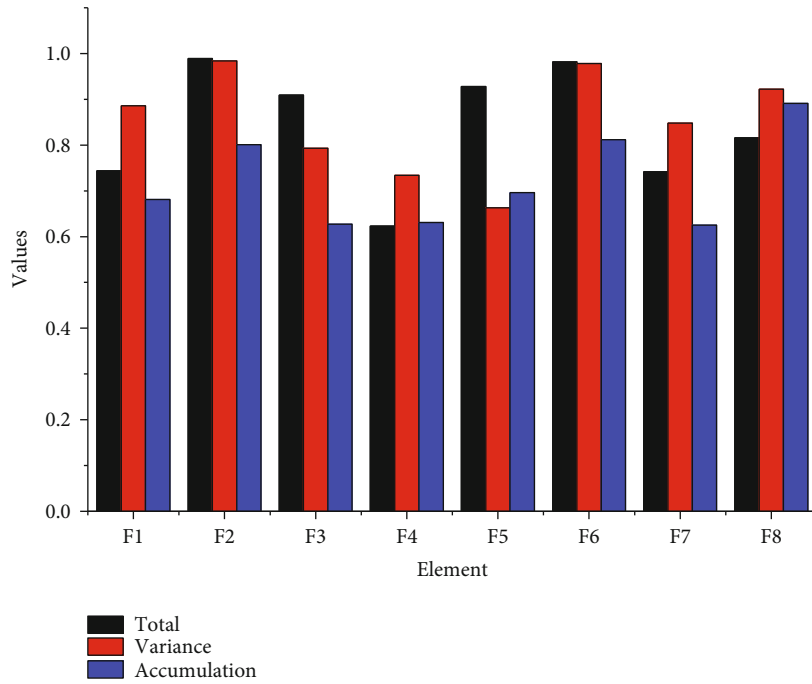


(a)

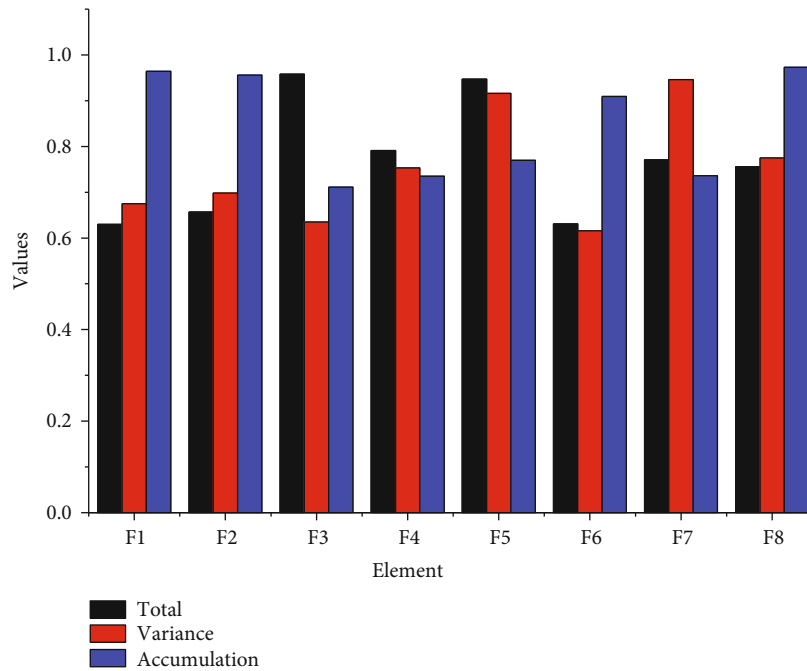


(b)

FIGURE 5: Continued.



(c)



(d)

FIGURE 5: The characteristic roots of the factors and the contribution of the total variance explained.

the shooting state as close as possible to the ideal value, combined with the analysis of technical movements; extract the free throw rhythm that is most suitable for the athlete; and abstract it into an arm trajectory curve. Using this curve as a template, repeated training can achieve muscle memory and improve the stability of free throws in the game. The first part targets the correction of the technical movements of the free throw shooter, targeting the detection of the four

steps of preparation, arm lift, squat, and extension in the first stage, and whether the technical features corresponding to each step are optimal. The second part is done by detecting the angle and speed of the ball's shot, through statistical analysis. Comparing the difference between this data and the theoretical minimum shot speed and the corresponding optimal shot angle, the athlete maintains the shot state as close to the ideal value as possible, and combined with the

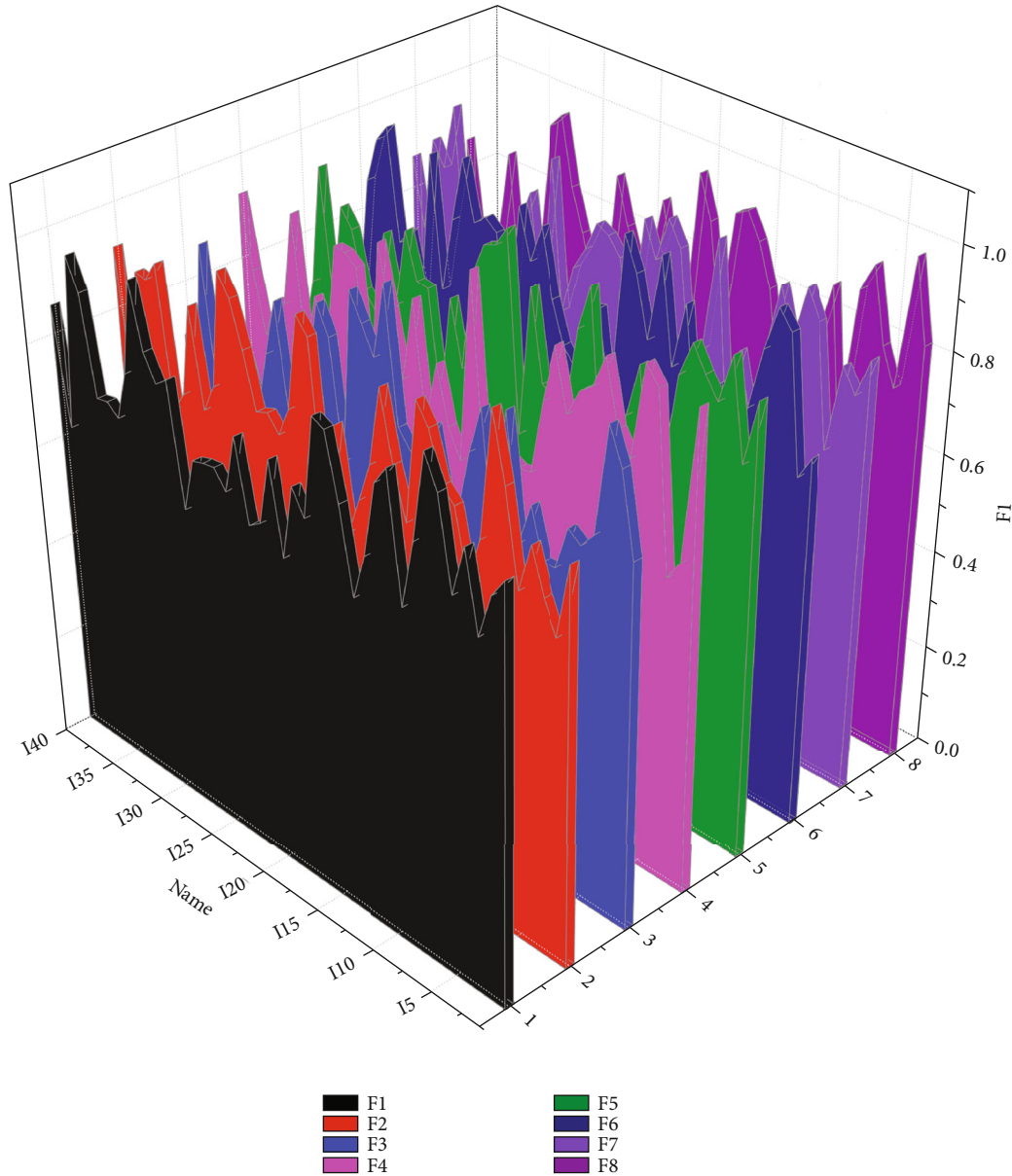


FIGURE 6: Factor loading statistics for each dimension.

analysis of technical movements, the most suitable free throw rhythm for the athlete is extracted and abstracted into an arm trajectory curve. Using the curve as a template, the repeated training achieves muscle memory and improves the stability of free throw shooting in the game.

The technical elements of each phase are described, and the process of wrist flicking on the ball during the shot is analyzed [21]. By approximating the trajectory of the ball as a parabolic motion subjected to gravity only, the relationship between the angle of the shot and the angle of entry can be derived based on physical knowledge, and the existence of the optimal angle of the shot and the minimum speed of the shot is argued.

The variables of the system are represented as a linear combination of a small number of common factors and special factors that affect only one variable. Factor analysis

refers to the study of extracting common factors from a population of variables by grouping several variables that are more closely related in the same category. Capturing these major factors can lead to a clearer analysis and interpretation of complex problems. First, the original group of variables is examined to determine whether it is suitable for factor analysis. (i.e., the original variables should have a strong correlation with each other). A common method for determining the number of factors is to determine the number of factors with the help of two criteria. The principal components with eigenvalues greater than or equal to 1 are selected as the initial factors, while those with eigenvalues less than 1 are discarded. And the number of factors is finally determined by combining with the gravel plot for validation. After that, the factors are further subjected to an effective naming interpretation.

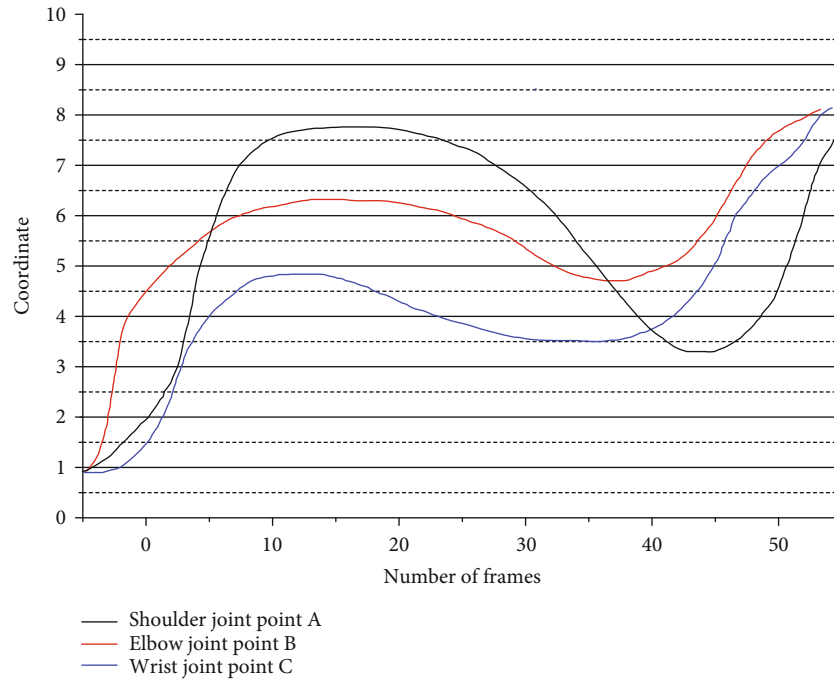


FIGURE 7: Longitudinal trajectory of arm joint points.

## 4. Analysis of Results

**4.1. Big Data Mining Performance Results.** Firstly, the k-s test was performed on the game data, the two-sided asymptotic significance of the test results for each group of values was greater than 0.1, and these data were considered to conform to a normal distribution. Secondly, the resultant data derived from the original algorithm were standardized by Z-scores, and the processing principles were as follows: in the statistics of the data, there were differences in the indicators, such as the number of hits and rebounds. Therefore, the differences in the units of indicators were eliminated; indicators with obvious functional relationships were excluded, for example, the total number of shots is the sum of the number of two-point shots and the number of three-point shots. Finally, 14 indicators such as i1, i2, i3, i4, i7, i10, i13, i17, i18, i20, and i35 were excluded; to facilitate in-depth analysis of team differentiation, the negative indicators were converted into positive indicators.

KMO and Bartlett's sphericity tests were performed on the data, and the KMO value of the data was 0.786 with a two-sided  $\text{sig} < 0.001$ . This indicates that common factors exist between the correlation matrices of the data clusters, and the data are suitable for factor analysis. The remaining 23 indicators were subjected to correlation matrix analysis to extract the common factors, and the results were rotated by the method of maximum variance to exclude the lower factor loadings ( $< 0.60$ ), resulting in the rotated factor loading matrix. And the principal components with eigenvalues less than 1 were excluded. The results are shown in the table; there are 7 factors with eigenvalues greater than 1, which are compatible with the theoretical concept. Their eigenvalues were 5.394, 3.842, 2.638, 2.353, 2.100, 1.730, and 1.506, with

a cumulative contribution rate of 72.455%, as shown in Figure 5.

Although factor analysis can reduce the influence of artificial subjective factors, the shortcoming is that it relies too much on sample indicator data and cannot reflect the preferences of decision-makers. In principle, the selection of question items not only is simply removed from the statistical strict requirements but also needs to be integrated according to the different research questions, the completeness of the meaning of the conception of the theoretical dimensions, and the significance level of each factor loading, high and low grouping, and other conditions under consideration. To ensure the applicability of the system factors, basketball experts and coaches were consulted to correct the categorization of indicators and determine the naming of factors according to the characteristics of basketball.

Its eigenvalues are 5.394, 3.842, 2.638, 2.353, 2.100, 1.730, and 1.506, and the cumulative contribution rate is 72.455%.

Only indicator i23 on the F6 principal component, for the number of steals, was loaded as 0.733. The ball-breaking skill is a very important basketball defensive skill. It is a comprehensive reflection of the team's defensive ability and the team's collaborative defensive formation. The i23 is more clearly defined in basketball theory and is close and highly correlated with the theoretical conception of F4 latitude (0.602). Therefore, i23 is classified as one of the F4 dimensions. The game of basketball requires not only technical and tactical cooperation but also perseverance, and high morale can be a factor in winning or losing the game. Among the many ways to boost morale—steals, caps, etc.—dunking is one of the most effective ways. The dunking technique has its unique characteristics: agitation, aggression, artistry, and

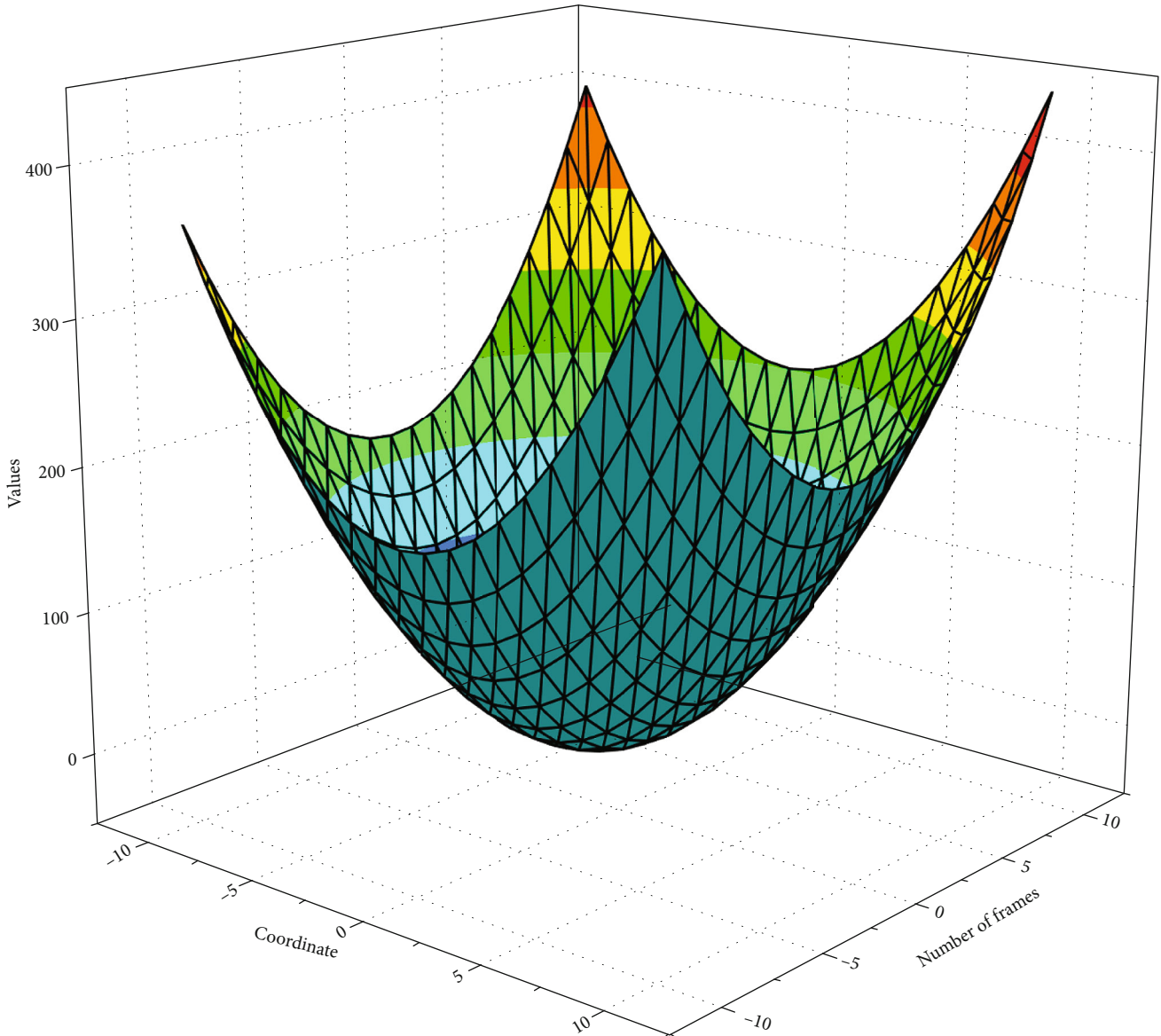


FIGURE 8: Template diagram of free throw extension trajectory.

skill. Expert opinions on this are divided. *i21* appears less frequently in Cuban games but plays a good role in boosting team morale and can be listed separately as an indicator of psychological aspects. However, there are more suggestions that dunks should be included as an offensive strength indicator. The subject is intended to be universal in the system and has a loading of 0.627 in the F1 latitude; therefore, *i21* is incorporated into F1.

This research is aimed at the universality of the system, and the load in the F1 latitude is 0.627; therefore, *i21* is incorporated into F1. These indicators are a better representation of a team's defensive skills and individual defensive skills. Therefore, it is defined as a defensive quality; the factors with high loadings in F5 are the number of three-point shots (0.647), two-point shots (0.838), and two-point tendencies (0.941). The two indicators of three-point tendency and two-point tendency can better evaluate a team's choice

of offensive style for the outside and inside. Therefore, this dimension was named an offensive choice. The factor loadings and common degree statistics of each dimension are shown in Figure 6.

In the study, the input video unit is firstly extracted in the foreground, based on which only the foreground pixels are traversed, which can improve the efficiency of the algorithm, and then, the corresponding RGB values are compared, and the pixel points that satisfy the marker color range are set to 255 and the other pixel points are set to 0 to obtain the feature binary map containing the marker points. To remove the possible irrelevant feature points, the corresponding erosion and expansion are performed. The binary map containing only the marker points is obtained by the above method. The contour of the binary map is then extracted to obtain the contour that contains each marker point, and the statistical analysis of the pixel



coordinates of the contour block is performed to obtain the statistical center of mass of the pixel block as the node coordinates.

**4.2. Simulation Results of Multiattribute Training Data.** The data used in the paper for the arm trajectory analysis were a basketball coach with standard skills and two average students, and the three could perform 80 sets of free throws each, with every 20 sets as a video unit. By performing image preprocessing algorithms on more than 15,000 frames of the three video units, to obtain the corresponding trajectory change characteristics, it can be found that there are respective periodic repetitions of the longitudinal coordinate trajectories of the three subjects' joint points, and Figure 7 is analyzed by plotting the individual free throw cycle trajectories of everyone.

Figure 7 shows the trajectory changes of the arm joint points of the two regular students, and it can be found that although there is no similarity in the trajectories as in Figure 7, the degree of similarity is still 75%, especially in the later trajectory transformation showing the same characteristics. By analyzing the frames corresponding to the maximum point in the trajectory change, it is found that in all the video samples, although the trajectories of each person are different, it can be found that the subject, as a basketball coach, can throw the ball at the highest point of his arm as much as possible due to his professionalism and the repeatability of his movements, i.e., the shot frames correspond to T frames; however, students A and B cannot do so, i.e., the distribution of the strike frames is from T-1 to T-3 frames. Therefore, to obtain the shot frames more accurately, a basketball localization detection algorithm is proposed to locate the coordinates of the center of the ball by improving the Hough transform. Then, we calculate the Euclidean distance  $D$  from the center of the ball to the wrist marker C and analyze its change curve to further analyze the characteristics of the shot frames.

We obtain the trajectory of the longitudinal coordinates of the arm joint A from T1 to T2 of the tested player and compare it with the technical model of that phase, and if the similarity reaches 60%, the player is considered to have a squatting step, and we further calculate the squatting depth and the minimum knee angle to judge whether the player's squat is in place or not and give the corresponding modifications. By obtaining the trajectory curve of the wrist joint C at the vertical coordinate from T2 to T3 during the extension process, the fluency of the extension is judged by comparing it with the standard extension curve. For example, Figure 8 shows the average trajectory graph normalized by the data of more than 500 hit free throws of athletes from multiple varsity men's basketball teams. As shown in Figure 8, the trajectory change from T2 to T3 is processed by normalization. The trajectory of the athlete's input data is matched with the trajectory of the T2 to T3 segment, and a match of 80% is considered as the rhythm of the stretching process meets the criteria; otherwise, it is considered that there is a lack of fluency in the stretching. The angle and speed of the stroke are compared with the optimal speed, angle, and angle between the arm and the horizontal plane corresponding to the height of the individual stroke to determine whether the stroke is reasonable.

We obtain the best shooting speed, the angle, and the angle between the arm and the horizontal plane corresponding to the shooting angle and speed with the individual shooting height and compare it to judge whether this shot is reasonable. Suggestions are given for "spinning the wrist too fast" or "spinning the wrist too slow" and "raise the arm appropriately to increase the angle to obtain a higher shot height." Suggestions are given for "too fast wrist flick" or "too slow wrist flick" and "raise the angle of the arm to obtain a higher strike height." Firstly, the algorithm is based on the automatic positioning of the shot frame. Secondly, a basketball detection and positioning algorithm is proposed using an improved Hough transformation algorithm to calculate hand speed status information. Finally, the technical actions are guided by the four steps of the free throw technique, and the technical trajectory and technical parameters of each step are analyzed to guide the athletes in their actions.

## 5. Conclusion

The basketball free throw-assisted training system applies computer, image processing, database, and other technologies to assist in physical training, which is becoming increasingly responsive to the times in today's intelligent and data-oriented world. This paper takes the recorded video of free throw shooting by the player as the original material and studies the technical steps and essentials steps for free throw shooting by analyzing the trajectory of the arm in the process of free throw shooting. By reviewing the relevant literature, the free throw technique is theoretically analyzed and summarized to deduce the existence of optimal shot speed and angle under the corresponding height. Then, the correctness of the theory was verified by analyzing a large amount of video data of free throw shooting by professional athletes and coaches, and the trajectory template of the arm in the free throw shooting process was extracted by digitizing the relevant technical characteristics through trajectory analysis. Through the actual movements of the athletes, combined with the theoretical analysis, the technical steps of free throw shooting are proposed: four steps of preparation, arm raising, squatting, and extension, and the time nodes of each step are obtained using the free throw shooting step time point acquisition algorithm. For the trajectory characteristics and technical parameters of arm raising, squatting, and extension, a large amount of statistical analysis was conducted to obtain the corresponding template trajectory. The corresponding technical guidance is given to the free throw shooter by comparing with the template trajectory in the assisted training system. At the same time, the technical parameters corresponding to the subject are calculated and compared with the standard parameters to achieve the purpose of assisting athletes' training.

## Data Availability

The data used to support the findings of this study are included within the article.

## Conflicts of Interest

All the authors do not have any possible conflicts of interest.

## References

- [1] P. Zuccolotto, M. Manisera, and M. Sandri, "Big data analytics for modeling scoring probability in basketball: the effect of shooting under high-pressure conditions," *International Journal of Sports Science & Coaching*, vol. 13, no. 4, pp. 569–589, 2018.
- [2] C. Yuan, Y. Yang, and Y. Liu, "Sports decision-making model based on data mining and neural network," *Neural Computing and Applications*, vol. 33, no. 9, pp. 3911–3924, 2021.
- [3] E. Morgulev, O. H. Azar, and R. Lidor, "Sports analytics and the big-data era," *International Journal of Data Science and Analytics*, vol. 5, no. 4, pp. 213–222, 2018.
- [4] R. Irfan, Z. Rehman, A. Abro, C. Chira, and W. Anwar, "Ontology learning in text mining for handling big data in healthcare systems," *Journal of Medical Imaging and Health Informatics*, vol. 9, no. 4, pp. 649–661, 2019.
- [5] P. Martínez-Santos and P. Renard, "Mapping groundwater potential through an ensemble of big data methods," *Groundwater*, vol. 58, no. 4, pp. 583–597, 2020.
- [6] H. Gong, N. M. Watanabe, B. P. Soebbing, M. T. Brown, and M. S. Nagel, "Do consumer perceptions of tanking impact attendance at National Basketball Association games? A sentiment analysis approach," *Journal of Sport Management*, vol. 35, no. 3, pp. 254–265, 2021.
- [7] G. Vinué, "A web application for interactive visualization of European basketball data," *Big Data*, vol. 8, no. 1, pp. 70–86, 2020.
- [8] F. Thabtah, L. Zhang, and N. Abdelhamid, "NBA game result prediction using feature analysis and machine learning," *Annals of Data Science*, vol. 6, no. 1, pp. 103–116, 2019.
- [9] A. Urbaczewski and R. Elmore, "Big data, efficient markets, and the end of daily fantasy sports as we know it?," *Big Data*, vol. 6, no. 4, pp. 239–247, 2018.
- [10] R. Metulini, "Filtering procedures for sensor data in basketball," *Statistica & Applicazioni*, vol. 15, no. 2, pp. 133–150, 2017.
- [11] H. Chen, "Neural network algorithm in predicting football match outcome based on player ability index," *Advances in Physical Education*, vol. 9, no. 4, pp. 215–222, 2019.
- [12] V. O. Kayhan and A. Watkins, "A data snapshot approach for making real-time predictions in basketball," *Big Data*, vol. 6, no. 2, pp. 96–112, 2018.
- [13] L. Li, Y. Zhao, and R. Nagarajan, "Optimising NBA player signing strategies based on practical constraints and statistics analytics," *International Journal of Big Data Intelligence*, vol. 6, no. 3/4, pp. 188–201, 2019.
- [14] J. S. Saltz and N. Dewar, "Data science ethical considerations: a systematic literature review and proposed project framework," *Ethics and Information Technology*, vol. 21, no. 3, pp. 197–208, 2019.
- [15] F. R. Goes, M. Kempe, L. A. Meerhoff, and K. A. P. M. Lemmink, "Not every pass can be an assist: a data-driven model to measure pass effectiveness in professional soccer matches," *Big Data*, vol. 7, no. 1, pp. 57–70, 2019.
- [16] B. Li and X. Xu, "Application of artificial intelligence in basketball sport," *Journal of Education, Health and Sport*, vol. 11, no. 7, pp. 54–67, 2021.
- [17] L. Chen, S. Gao, and X. Cao, "Research on real-time outlier detection over big data streams," *International Journal of Computers and Applications*, vol. 42, no. 1, pp. 93–101, 2020.
- [18] M. Migliorati, "Detecting drivers of basketball successful games: an exploratory study with machine learning algorithms," *Electronic Journal of Applied Statistical Analysis*, vol. 13, no. 2, pp. 454–473, 2020.
- [19] U. Dick and U. Brefeld, "Learning to rate player positioning in soccer," *Big Data*, vol. 7, no. 1, pp. 71–82, 2019.
- [20] M. Carpita and S. Golia, "Discovering associations between players' performance indicators and matches' results in the European soccer leagues," *Journal of Applied Statistics*, vol. 48, no. 9, pp. 1696–1711, 2021.
- [21] X. Zhang and J. Sun, "Discussion on new media communication strategy of sports events based on large data technology," *Cluster Computing*, vol. 22, Supplement 2, pp. 3395–3403, 2019.

## Research Article

# Integrated Design of Graduate Education Information System of Universities in Digital Campus Environment

Jing Ma <sup>1</sup> and Bo Feng<sup>2</sup>

<sup>1</sup>Department of Postgraduate, Hangzhou Normal University, Hangzhou 311121, China

<sup>2</sup>College of Electrical & Information Engineering, Shaanxi University of Science & Technology, Xi'an 710021, China

Correspondence should be addressed to Jing Ma; [majing2012hsd@163.com](mailto:majing2012hsd@163.com)

Received 7 August 2021; Revised 31 August 2021; Accepted 1 September 2021; Published 30 September 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Jing Ma and Bo Feng. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This study takes the digital campus construction planning of the high school as an example and determines the requirements of the postgraduate management information system under the digital campus environment through the analysis of the overall framework and technology of the digital campus. Combining the current situation of computer technology, network technology, and the actual situation of our university, the current mainstream B/S three-layer architecture is adopted, the web adopts the current popular Java Server Pages technology, and the struts framework connects to the Oracle backend database through the Java Database Connectivity interface to design the browser-side and server-side programs. The struts framework connects to the Oracle backend database through the Java Database Connectivity interface to design browser-side and server-side programs. The functional model and data flow model of the system were established through a detailed and effective analysis of the entire workflow of postgraduate students' training management during their school years. Then, the system analysis, design, and drawing of the swim lane diagram and data business flow diagram were carried out. The system was designed in detail in terms of system architecture, development tools, functional modules, and database design, and the core module of training program making in postgraduate training management was highlighted as an example to discuss the principles and methods in the construction of departmental business systems and informatization under the digital campus environment, and a flexible and efficient postgraduate management information system was realized. It standardizes the construction of data standardization in universities; does a good job of standardizing and normalizing information; improves the accuracy, validity, and real-time production of data collection and the real and safe unified management of historical data; and provides scientific and reasonable data support for the leadership to make relevant decisions.

## 1. Introduction

With the deepening of education reform, the scale of postgraduate enrollment has been expanding, and the data to be recorded and processed in the training management work has increased exponentially. To standardize management and improve the level and efficiency of postgraduate training management, many universities have established postgraduate management systems and related teaching support platforms one after another to realize the collection, processing, and statistical functions of basic teaching information, which reduces the manual processing work of management personnel, improves the working efficiency, and brings great convenience to teachers

and students of the whole university [1]. It reduces the manual work of management staff, improves work efficiency, and brings great convenience to the teachers and students of the school. However, from the perspective of the development trend of graduate education and teaching, there are still some unsuitable places in the informatization of graduate students in colleges and universities. However, from the perspective of the development trend of postgraduate education and teaching, there are still some unsuitable places for postgraduate informatization in colleges and universities, mainly in the following aspects: many postgraduate systems in colleges and universities were built earlier; the system construction did not follow the national standards and industry standards; the data structure

and data content were not standardized; it is difficult to ensure the consistency of data, resulting in difficulties in information and data sharing between systems; and it is difficult to achieve an upgrade and maintenance of the system [2]. Many schools do not include the construction of graduate education informatization into a purposeful, planned, and step-by-step system engineering project, and even if there is a better hardware and software environment, it is difficult to mobilize most teachers and students to participate in the construction of education informatization to the greatest extent due to the lack of a corresponding standardized system and standardized management [3]. Many systems are designed from the needs of administrators, only designing management functions, with less support for cultivation management closely related to students' learning activities, which cannot support the whole process of graduate students' learning activities, making it difficult for administrators to track and monitor the completion and quality of graduate students' studies and ensure the smooth progress of research work [4]. After collecting a lot of teaching information into the system, it only does simple query and statistical functions without deep processing and lacks multidimensional statistical analysis functions, which makes it difficult to provide management decision analysis information to school leaders and managers.

This study intends to draw on the development experience of other perfect systems, compare the more popular system development technologies, analyze the specific needs of the postgraduate training process in the light of the overall construction of our university digital campus, and share the data information of other systems in the digital campus through the data exchange and sharing platform. For example, the personnel information in the personnel system, the financial information in the financial system, the undergraduate major course information in the teaching service system, and the student accommodation information in the coordination management system, etc., to design a set of management information system suitable for the actual situation of postgraduate management. However, due to the lack of a corresponding standardized system and standardized management, it is difficult to mobilize teachers and students to devote themselves to the construction of educational information. Many system designs are based on the needs of managers, only designing management functions, and there is little support for training management that is closely related to student learning activities. It integrates various heterogeneous data in the digital campus, unifies the management of semistructured and structured data, eliminates information silos, and realizes real-time data sharing among systems [5]. For example, it provides authoritative data such as basic information of graduate students, cultivation, and degrees for the data sharing center; it extracts basic information of faculty personnel and additional information such as payment and accommodation of graduate students from the data sharing center, which ensures the consistency of data [6]. Before the design and architecture of the system, a prospective study on the data standards of the university was conducted, and the coding rules of human, financial, and material aspects were planned uniformly from the university, which laid a good foundation for the construction of the data sharing center in our university. In the design pro-

cess of the postgraduate management information system, it is designed to exchange data from the digital campus shared data platform and reserve the data interface for public information calling from the university's portal or other systems calling. For example, the integrated coding rules for personnel information from students to faculty members can provide a reference for the development of subsequent systems by eliminating data cleaning, transcoding, and filtering for the personnel database in the later stage.

It proposes the J2EE development platform of the academic affairs network management system; introduces in detail the digitalized resource construction, student academic supervision, and effective control, assessment of teachers' teaching and students' learning, and student file management modules of the university; and proposes how to optimize the functions of the information-based academic affairs network management system for the paper focuses on the digital campus teaching and learning management of universities. This paper focuses on the design and implementation of the digital campus academic affairs management system of colleges and universities and analyzes and discusses each functional module in the digital campus academic affairs management system of colleges and universities, which has a certain innovative value and has certain reference significance for the development of the digital campus academic affairs management system of other colleges and universities. In addition, computers can be used to select elective subjects for college teaching and carry out teaching assessments, which greatly saves the work pressure and work content of administrative teaching staff. Computer network technology allows each student to choose elective subjects independently through the computer, which provides sufficient autonomy for students and greatly improves work efficiency, allowing teaching staff to have more time to deal with other things and meet the needs of students at different stages.

## 2. Current Status of Research

Although the current university networked teaching management model has improved, it still lacks a lot of rationality and humanized design. Based on the above considerations, the development of an artificial intelligence-driven teaching management information system suitable for various needs is particularly important for improving the efficiency and quality of managers and enhancing the image of the school [7]. For example, the British scholar Aithal and others mentioned that some institutions in the UK and France have built information technology earlier, and each university focuses on the construction of a web platform in the academic affairs network management system [8]. Through decades of development, they already have a perfect and mature teaching system, teaching service system, office system, library system, campus entertainment system, and so on [9]. These system platforms have had a great impact on students, teachers, administrators, and community residents, facilitating students' spare time, improving office efficiency, improving learning environment and methods, and accelerating the pace of transformation of an information-based

society [10]. The digital campus is better than UEAS Academic Affairs Management System, which combines the advantages of both and builds a better application system, deploying C/S computing mode for parts that need a lot of calculations, such as data import and data analysis, and B/S computing mode for parts that do not need too many calculations, such as information query and display [11]. The C/S computing mode is open to a few users who need it, and the B/S computing mode is developed for all users of the system [12].

Many schools have established student information management systems, which provide a solid foundation for modern network education. The technical aspect of the management information system led to innovation and reform of the traditional office model, thus creating a previously unavailable, new form of efficient modern office model [13]. As we all know, for the long-term sustainable development of each enterprise, there is a need for an efficient and convenient management system; they will mostly choose reliable public generally recognized software companies to cooperate in the development, but the required development costs and maintenance costs will become very difficult to bear, which for the market users, will be a cost-effective automated management information system [14]. This paper will design and study the most cost-effective management information system to meet this demand [15]. In the initial stage, the construction of information technology in universities only stopped at the registration and storage of data and only established a relatively closed local network on campus, and the individuals who used this network only used it in the local area network, and with the improvement of requirements, the university built a medium-sized campus network, which improved the efficiency of using the campus network and accelerated the improvement of teaching quality to a certain extent [16]. A large information center was built, and from the initial embryonic stage to the present day, a complete information system has been built, which has greatly improved the degree of digitalization of the school [17]. From the failed cases, we summarized where its shortcomings lie and developed a student management system suitable for education according to our characteristics [18]. At present, all major universities have built and improved their campus networks, and the investment in hardware has taken shape, slowly entering the middle stage of the information technology era.

Based on ensuring information security, it is registered and published on the school's unified application platform to realize the sharing of information results within the school, and on this basis, it is integrated with Internet applications. The supply of highly experienced application services and integrated access to services to students greatly increases the dependability of users. With integrated services centered on high usage, campus faculty and students can not only enjoy the services provided by the integrated service platform but also provide their own opinions on problematic campus facilities or campus services that need to be improved, supporting decision-makers at all levels to improve and study business models. The test cases of some functions of the digital campus teaching management sys-

tem are explained. The key involves book lending test cases, user login test cases, grade retrieval test cases, adding academic records test cases, and specific descriptions of the system performance tests.

### 3. Analysis of Integrated Design of Information System for Graduate Education in Digital Campus Universities

*3.1. Technical Analysis of Graduate Student Education Information System.* The system uses an on-demand B/S technology architecture. This system also uses a J2EE-based development framework and runtime environment to maximize the inclusion and integration of existing and to-be-built applications [19]. For example, provide the data sharing center with authoritative data such as basic information, training, and degrees of graduate students; extract basic information about teachers and workers from the data sharing center; and additional information such as postgraduate payment and accommodation to ensure the consistency of the data. The SOA service-oriented architecture simplifies the processing mode of business process integration between systems, simplifies the difficulty of cross-system interface integration due to changes in business requirements, and achieves loose coupling between systems [19]. The Oracle database with excellent performance is selected as powerful. A distributed software architecture is used to realize the integration of data, services, and business processes utilizing application integration, and the digital campus is built in steps under a unified top-level design. Each system operates independently, and cross-system invocation and collaboration are realized through standard and clear interfaces.

The server is usually a high-performance PC, workstation, or minicomputer with a large database system such as Oracle or SQL Server. The B/S architecture has three main components: client browser, application publishing server, and database server, as shown in Figure 1.

The construction of a digital campus is a long-term and continuous construction process that matches the strategy of the university. It is necessary to consider the results and data of the university's existing information technology construction, and at the same time, it is necessary to lay the foundation and indicate the direction for the future development of the university's information technology, so that the construction of the digital campus is a continuous improvement and continuous updating process over time and can effectively form the long-term accumulation of the university [20]. It does not break with the passage of time, changes in technology, changes in management personnel, and changes in thinking. Digital campus construction is to provide advanced and comprehensive informatization solutions for universities to build their core competitiveness through internal resource integration and accumulation of existing informatization construction achievements. Digital resource construction, student academic supervision and effective control, teacher teaching and student learning evaluation, student file management, and other modules propose how to optimize the functions of the information-based

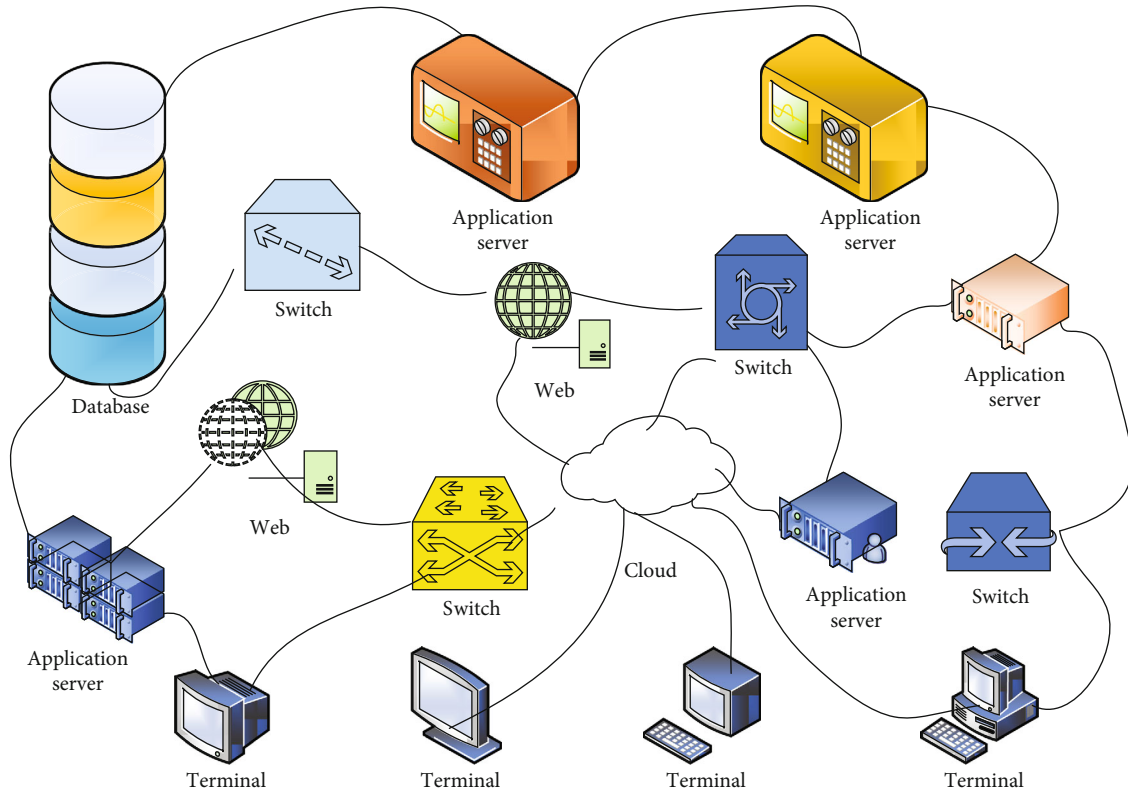


FIGURE 1: B/S architecture diagram.

educational administration network management system and provide for the sustainable development of university educational administration network management reference effect. It guides the construction of a digital campus by formulating a long-term informatization construction plan of the university, comprehensively constructing university-level management information system, optimizing existing management process and management mechanisms, improving administrative management efficiency, and saving management cost of the university while ensuring the realization of strategic objectives of the university. Strengthen the construction of a basic support environment to ensure the safe, stable, and reliable operation of the system; build the basic platform of digital campus based on the construction framework of "hardware cluster, data concentration, application integration, and service integration." Optimize the means of collaboration among personnel at all levels of colleges and universities, create a life service platform, teaching service platform, and scientific research service platform, and build a comprehensive digital campus service platform for teachers, students, and staff [21]. It standardizes the construction of data standardization in colleges and universities; standardizes and unifies the information; improves the accuracy, validity, and real-time of production data collection and the real and safe unified management of historical data; and provides scientific and reasonable data support for the leadership to make relevant decisions.

In the era of information technology, the education management of colleges and universities is also facing greater challenges. The construction and management of digital campuses

in colleges and universities are the primary factor affecting the continuous improvement of the quality of education and teaching in colleges and universities [22]. In Huizhou Engineering Vocational College, the college proposes to pay attention to the construction of a digital campus network system, and in the teaching-centered mode, it is necessary to guarantee good teaching quality and good teaching mechanism, precisely for the establishment of the current higher vocational college at the early stage, to establish an integrated working system platform, which has more efficient work efficiency compared with the traditional campus management and can also serve the teaching staff and students accurately [23]. It allows data to generate more shared value and establishes a data intervisiting mechanism. Focusing on on-campus information applications, it solves various current management problems and serves campus faculty and students. This paper focuses on the design and implementation of the digital campus academic affairs management system of colleges and universities and analyzes and discusses each functional module in the digital campus academic affairs management system of colleges and universities, which has a certain innovative value.

Many colleges and universities implement the management mode of large academic affairs and put all the work related to education and teaching into academic affairs, so whether the school's academic affairs management is smooth and the quality of academic affairs management is directly related to the quality of education and teaching. In recent years, Huizhou Engineering Vocational College has been upgraded to a higher vocational college by merging several secondary schools, and the management mode needs

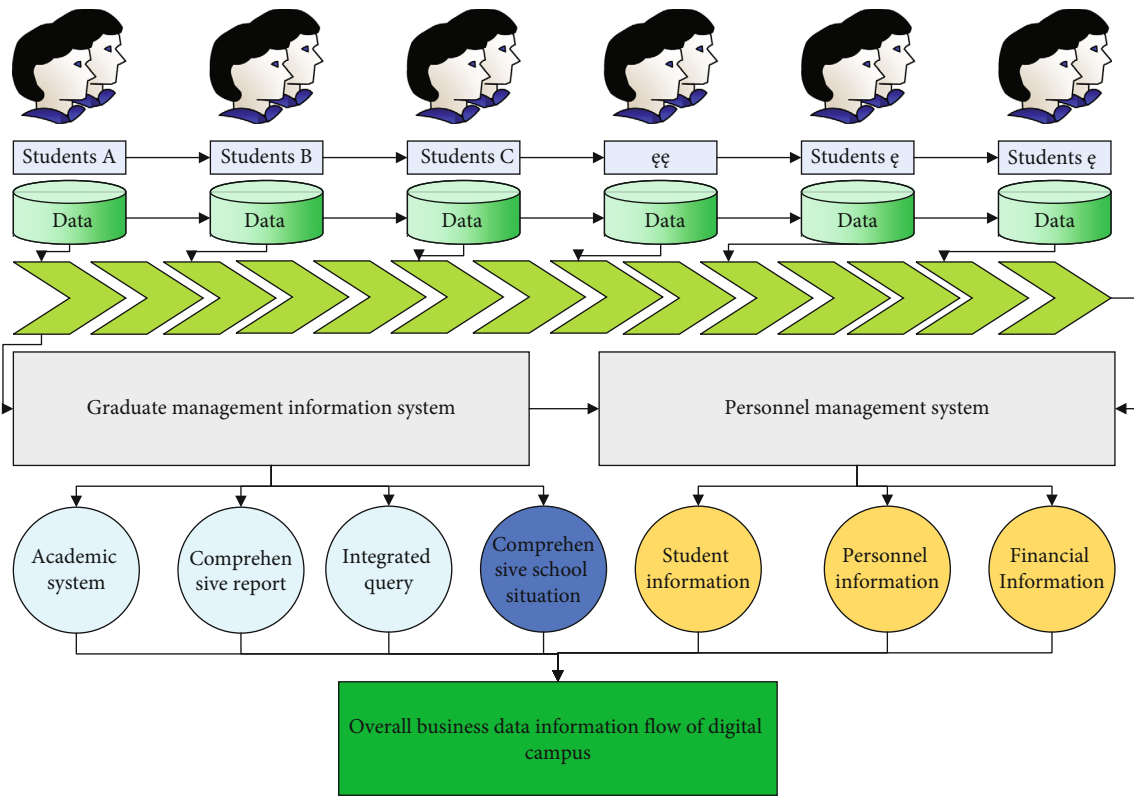


FIGURE 2: Overall business data information flow of digital campus.

to be changed continuously. At the same time, as the college continues to expand and transform, the academic administration will also face increased pressure. To maximize the inclusion and integration of existing and to-be-built applications, through the SOA service-oriented architecture, it simplifies the processing mode of business process integration between systems, simplifies the difficulty of cross-system interface integration due to changes in business requirements, and realizes loose coupling between systems. To adapt to the changing needs of academic affairs management, the traditional manual work can no longer meet the needs of complex academic affairs information management, which requires us to change our concept and adopt modern means to manage the current academic affairs information and continuously improve the management of academic affairs management for all teachers and students of the school, as shown in Figure 2.

The construction of a digital campus requires the school to establish its own data code rules, and to the maximum extent consistent with the education management informatization standard (Ministry of Education), relevant national standards, and relevant industry-standard codes, reduce the duplication of code usage, and consider the need for future code expansion for the construction and maintenance of code rules, i.e., school-defined data standard rules. The database server adopts a minicomputer system, and through physical partitioning technology, the informatization campus database and the one-card database are run separately in different partitions, while a separate partition is used as the environment for database testing and development. It

is planned to add a set of disk array systems with the same configuration as the main storage system in the offsite server room as the offsite disaster recovery center and realize the synchronous or asynchronous partial or complete replication of data from the main storage center to the disaster recovery center through the configuration of fiber optic switch plus long-distance transmission interface module and data mirroring software.

The application servers are recommended to use WebSphere distributed load balancing deployment architecture and the entire deployment. The outermost layer is a dedicated hardware load-balancing device and can be deployed as a dual-computer active-active high availability mode, which is responsible for providing load balancing and high availability guarantee for HTTP (80) traffic from users to the HTTP front-end server; the middle is two WebSphere HTTP servers, which act as task schedulers for the back-end. These two HTTP front-end servers function as task schedulers of the information portal platform and digital campus application servers at the back-end and are responsible for providing load balancing and high availability for user requests to access the information portal and application systems at the back-end; the back-end is a WebSphere Application Server Cluster member server. The construction of the digital campus is a process of continuous improvement and renewal over time, which can effectively form the long-term accumulation of the school. It does not lead to a break with the passage of time, technological development and changes, changes in management personnel, and changes in thinking.

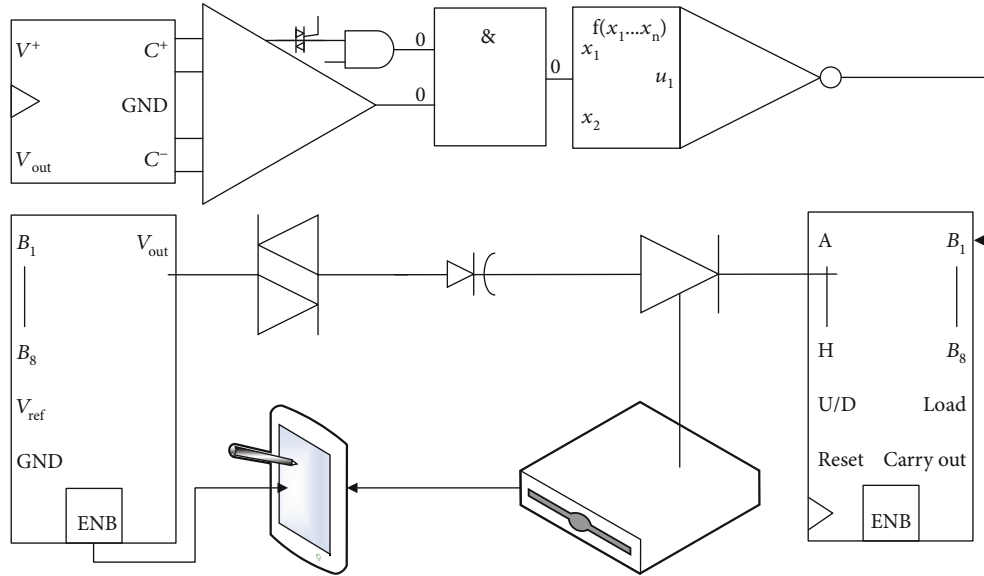


FIGURE 3: Schematic diagram of the unified login and authentication process.

The graduate school and colleges determine which courses need to be offered each semester according to the cultivation program and calculate the number of required and elective courses according to the student cultivation program and provide comprehensive inquiries of course information, output, and print course statistics, and other functions. The graduate school and all colleges adopt the computer-aided manual scheduling method according to the courses already offered in the current semester to determine the teachers, time, and location of each course, and the system provides strict conflict verification of teachers and classroom resources and provides automatic generation and query functions of the curriculum. Students can be scheduled manually during the scheduling process, and the system will automatically check whether students have conflicting class times. When scheduling classes in a new semester, the training office or college can refer to previous scheduling information and copy the scheduling information of certain courses as needed to reduce the maintenance workload.

**3.2. Digital Campus Environment Living System.** The smart campus software system is the window of the school's informatization construction, which can show the information of each interface for students and can allow users to easily find the desired information or resource portal, etc, reduce code reuse rate, consider the need for future code expansion, and carry out the construction and maintenance of code rules, that is, calibrate data standard rules. The database server adopts a minicomputer system and runs the information campus database in different partitions through physical partition technology. All the platform systems of the smart campus should be reflected in the school portal, and school news and notices can be released in time to show the development and progress of the campus. The construction of the campus information portal is not only a window for the school to display to the outside world but also an

important platform for teachers, students, faculty members, and other personnel to carry out unified identity authentication, based on which the entrance to other application systems can be easily found, providing great convenience to users. We believe that an important symbol of a digital campus is the construction of a campus information portal [24]. If the campus information portal is not well planned, it will be difficult to carry out the whole digital campus construction in depth. The training office or college can refer to the past scheduling information and copy the scheduling information of certain courses as needed to reduce the maintenance workload. The information portal is a center of various information collections, which can provide relevant information inquiry and release some information related to the column, which will be displayed on the school's portal after review. The campus portal is a large network platform integrating news release, content management, video on demand, image retrieval, full-text search, and other functions, as shown in Figure 3.

An intelligent access control system uses a card access control machine, license plate recognition, remote identification, and other access control equipment for linkage integration. In addition to controlling the entry and exit of vehicles, it can further calculate and control the number of live-restricted parking spaces, strengthen the antitheft/prevention function, and make the vehicles pass through the entrances and exits more effectively identified and managed [25]. All-important departments of the college, training areas, and machine rooms are accessed using access cards, and the permissions of each faculty member and student can be managed in the background. On this basis, it can be combined with a personnel attendance system to realize intelligent attendance and unified access to the college network for remote management. The digital campus is a system that takes the network and digital information as the basis and makes the teaching resources more fully utilized



by students, teachers, and managers. Teachers, in addition to having the same operation rights as students, can also access the teaching videos and animations of relevant courses through this system. Administrators can use this system to efficiently integrate school-related information and can release daily school arrangements to students and teachers through this platform, such as notification of school closures and power outages, so that the information notified to students, teachers, and other administrators can be notified within the first time. The above are the functions that can be provided by the smart card, but some additional functions will be added later to improve the interface of the system, to improve the quality of the system to achieve the purpose of improving the core competitiveness.

It completes the functions of the electronic library card and becomes the input medium of the library management application system. The system replaces the campus card with the library card. Students need to verify their identity when entering the library, and they need to swipe their cards at the card machine at the entrance of the library before the library will open, and they also need to swipe their cards to register their identity information when borrowing books. Through this system, the relevant information of the school can be efficiently integrated, and the daily arrangements of the school can be released to students and teachers through this platform, such as the notice of rest and power outage, so that the information notified by the school can be notified in the first-time students, teachers, and other administrators. The system uses the student card instead of the key, when students enter the dormitory need to brush the card, identity verification, only in the corresponding dormitory building student card can open the dormitory door, in the dormitory such as the population, male and female dormitory main channel, training building entrance to install access control equipment, access control system is conducive to improving the security of the school, prevent irrelevant workers step into the campus, affect the normal campus environment order, effectively prevent internal theft, burglary, and campus violations from occurring. The consumption system mainly includes the meal charge system for teachers and students, the supermarket charge system, and the drinking water charge system. The main function of the consumption system is to make the daily consumption of students and teachers inside the school safer and more transparent, for example, eating in the school's dining hall will send the daily consumption records to the APP, which can show the dining consumption for a month and the average daily consumption, etc. In addition, other consumption of students inside the school will also be recorded, such as drinking water and supermarkets, in which no cash consumption and old card spending.

Teachers and students will be integrated with the comprehensive application system to realize data sharing so that the comprehensive application system for teachers and students will become a whole and create a convenient, fast, and humanized campus electronic life for teachers and students. It is a comprehensive information platform for teachers and students to enroll, teach, train, live, seek jobs, learn skills, graduate, and resign. Nowadays, in the era of

rapid development of information technology, this is both an opportunity and a challenge for the school management system. The establishment of an integrated work system platform has more efficient work efficiency than traditional campus management and can also accurately serve faculty, staff, and students, to a greater extent, allow data to generate more shared value and establish a data exchange mechanism. On the one hand, the background of this era provides a more excellent network environment for the information system, and on the other hand, the information system faces more dangers due to the rapidly developing network environment. While the business is expanding, the emergence of education administration system not only provides users with a convenient and quick way to manage various tedious education affairs but also liberates the education department, streamlining administration, reducing recurring expenses, and using existing resources can bring greater benefits, as shown in Figure 4.

Nowadays, in the era of rapid development of information technology, this is both an opportunity and a challenge for the school management system. On the one hand, the background of this great era provides a more excellent network environment for the information system, and on the other hand, the rapidly developing network environment makes the information system face more dangers. While the business is expanding, the emergence of the education administration system not only provides users with a convenient and quick way to manage various tedious education affairs but also liberates the education department, streamlining administration, reducing recurring expenses, and using existing resources can bring greater benefits. The educational management system has been developed following the characteristics of universities. The key covers teaching management, teacher management, curriculum management, examination management, and training management [26]. The rational arrangement of these divisions of labor makes staff members more focused on their jobs and greatly improves the efficiency of their work.

To realize the decentralized campus monitoring system, we must first ensure the stability and security of the database and secondly realize the perfect unification of the database and each technology, such as network monitoring [27] and authority allocation, and unified management of all devices and users. The system can greatly simplify and standardize this process, realize information sharing and business collaboration between the upper and lower departments, and connect with departmental budgeting, recruitment, accounting, and fund clearing systems. The establishment of an online payment system mainly solves the problem of reimbursement of research funds. In colleges and universities, teachers conduct various kinds of research, in which the accounting and issuance of funds generated is a complex process. This system can greatly make this process streamlined and standardized; realize information sharing and business collaboration between upper- and lower-level departments; connect with a departmental budget, procurement, accounting processing, and fund clearing systems; realize asset change application; approve financial accounting; and realize the whole life cycle management of the whole assets.

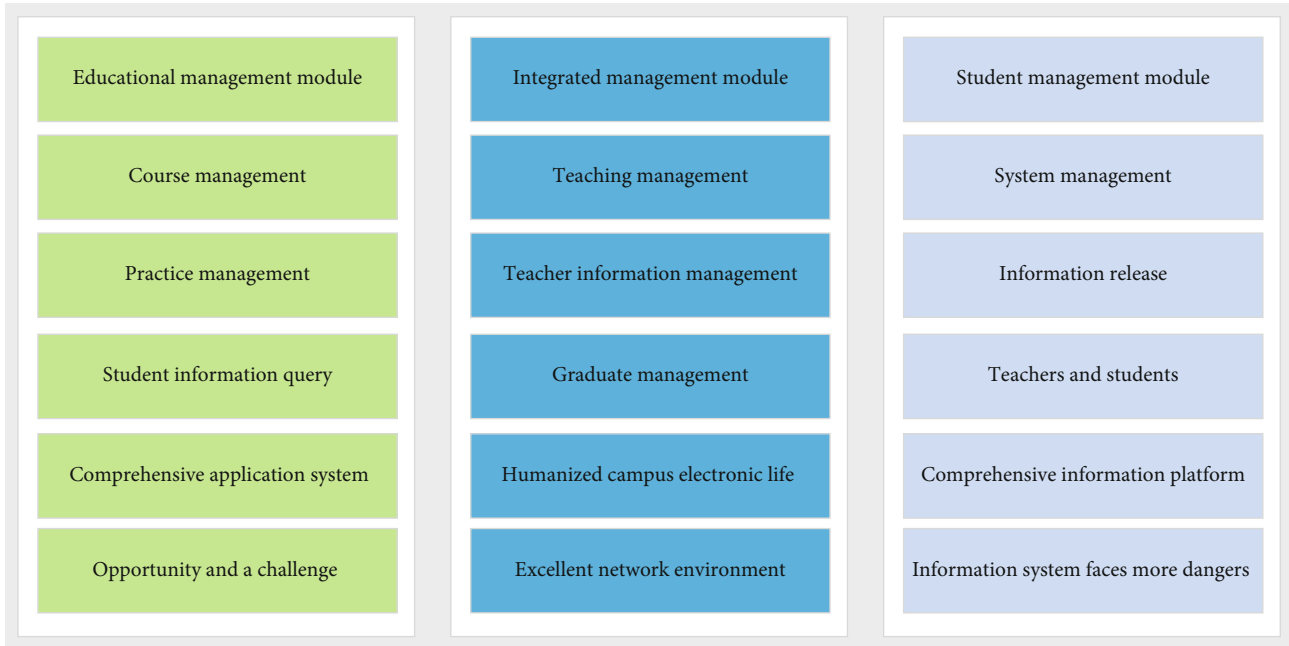


FIGURE 4: Management system function module.

## 4. Analysis of Results

**4.1. All-in-One System Testing.** Software testing, that is, the use of test equipment to carry out performance and functional tests on the product according to the test strategy and procedures, and even the design of different test equipment according to the requirements, was designed to maintain the system and the comprehensive evaluation of the various types of problems that arise in the test strategy. After the test cases executed, the exceptions are followed up to ensure that these developed products meet the detailed regulations. When designing the above test cases, the status of illegal user input and boundary conditions should be analyzed. The user's software should be tested to prevent any surprises. Confirmation of the test results should be carried out, and the errors of the test are often determined by another person when carrying out the confirmation. For the more prominent problems, a meeting can be considered to discuss and determine. With the help of the test results, it is possible to find out if the problem exists and the severity of the problem. The design goal of the test case for adding student status is to test whether the function of adding student status can be used to verify whether the user's input information is legal. When the data entered by the user is incorrect, the root cause of the error should be given and entered again. The goal of the Add Student Registration test case is to test whether the Add Student Registration feature is available to verify that the user's entry is legitimate. When the data entered by the user is wrong, the root cause of the error is given and entered again; after the user enters the accurate test data, is it possible to add the school registration information. During the system testing period, three rounds of testing were implemented, and the results of each round were counted after finishing the testing operation, and the

related fixes were implemented after each round of testing. The bug count record table for the three rounds of testing is detailed in Figure 5.

Performance testing intends to verify whether the performance of the application system can meet the relevant regulations of users and can learn the outstanding problems such as system bottlenecks and performance in the application system and provide powerful help for software improvement and overall optimization. In the teacher space, teacher users can complete three basic tasks: managing student information, managing course videos, and live teaching. In student information management, teachers can view the information of all students who choose the course they teach, as well as manage students' grades after completing the course. In Manage Course Videos, teachers can view their existing course video information and add new course videos. Once teachers upload their courses, they can see the courses they have added on the Watch Course page. Figure 6 shows the interface for watching course videos, where teachers can see their course type, number, name, and view the entire list of course videos. In addition, teachers can enter under the Search Course Videos tab to search for other teachers' course videos, which can also be watched and studied after purchase.

Software testers are generally separated from system developers, i.e., software testers are specialized in software testing. Black box testing is mainly functional testing. It mainly tests the functions of the software. It is assumed that the tester does not understand the functional structure of the system and knows the program structure of the system; the tester mainly tests whether the various functions of the software system meet the requirements. White box testing generally uses the test case approach, in which the system functions of the software are visible to the tester and the software tester understands the functional structure of the

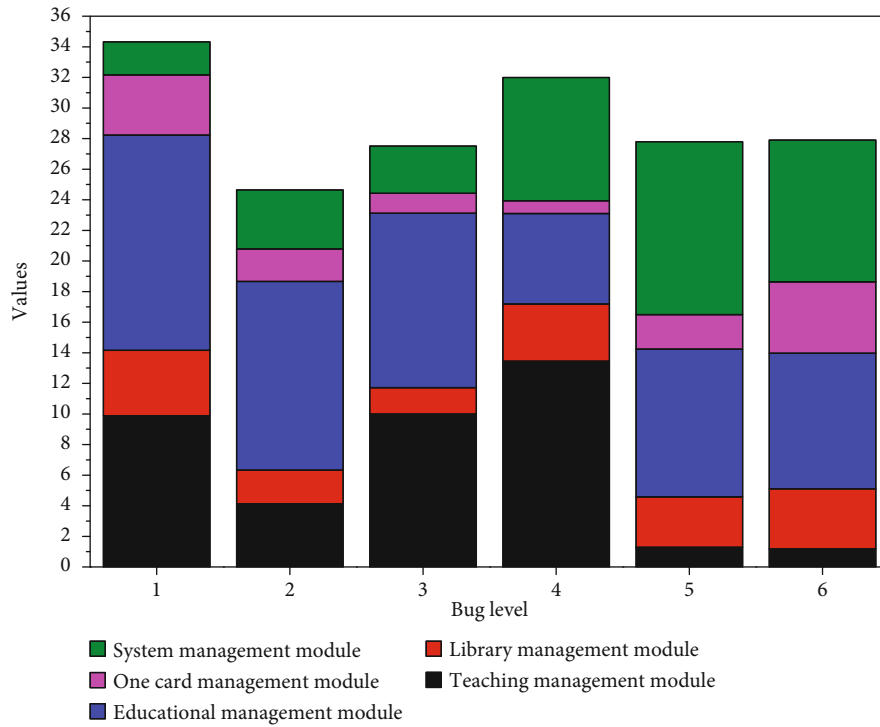


FIGURE 5: User login test case.

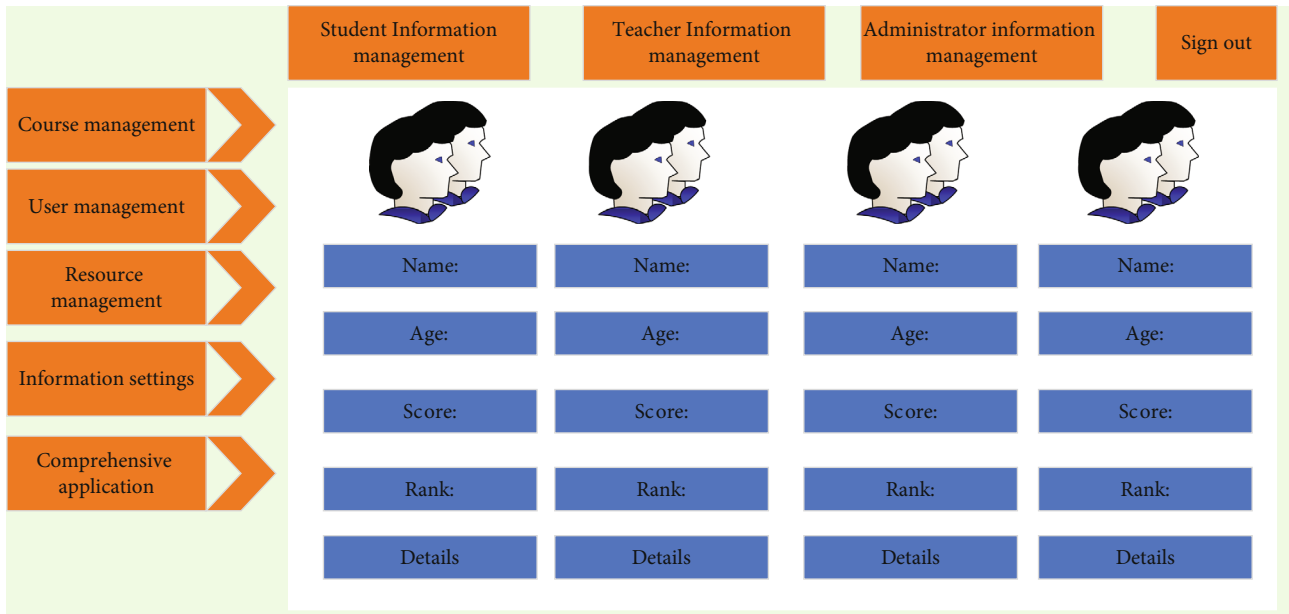


FIGURE 6: User information management.

system and knows the program structure of the system; black box testing, which is mainly functional testing, mainly tests the functions of the software, if the tester does not understand the functional structure of the system and knows the program structure of the system. The tester does not understand the functional structure of the system and knows the program structure of the system; the tester mainly tests whether each function of the software system meets the requirements.

4.2. Performance Test Results. In addition to completing the basic functions, the load capacity of the system must be tested to make the user experience better. In this paper, we use HBase, a nonrelational database with massive resource storage, to form a high-performance cloud storage environment during the development of the system, and we conduct a stress test on the database below. First, we import millions of simulated data in the educational storage resource table and then conduct four tests of sequential read, sequential

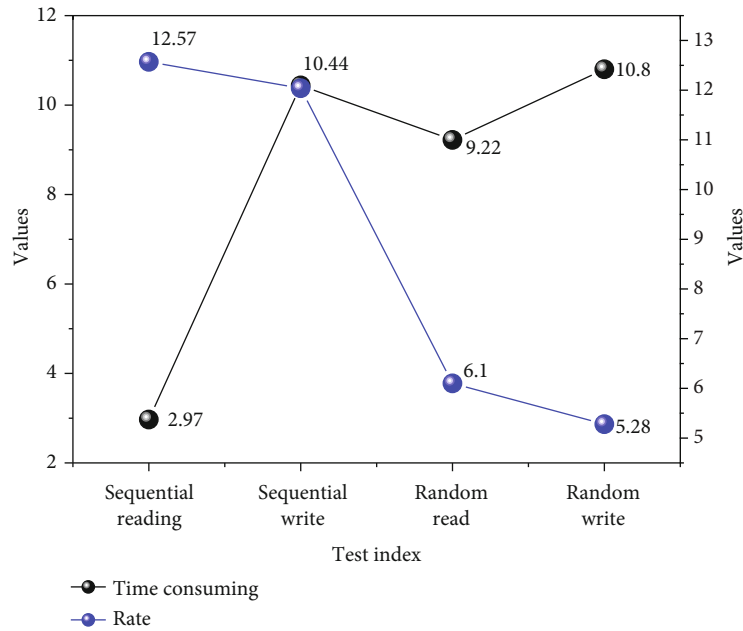


FIGURE 7: Database stress test.

write, random read, and random write by the test tool that comes with HBase. The test results are shown in Figure 7.

From Figure 7, using the HBase database as the cloud service center of the system can complete the storage of massive educational resources and meet the access requirements of users. About the relevant requirements of digital campus construction, a unified information portal platform design, a unified identity demonstration platform design, and a unified information resource planning quasicomstruction are given; the above design enables the campus to integrate and extend the existing system, which greatly reduces the cost and achieves the expected service effect. In the construction of the digital campus teaching management system of higher education institutions, various information resources of the school are integrated in a scientific and standardized way, and unified authority control, resource management, and user management are formed so that the school can develop better with the reengineering of organization and business process; system innovation and management innovation are carried out through the construction of the project, and finally, information education, scientific management, and choice are completed in a more standardized way. The basic functions of the digital campus teaching management system are realized, and on this basis, through the exploration of the construction of digital campus teaching management system in higher education institutions, and employment-oriented digital campus structure is formed, combining the characteristics of higher education institutions, with comprehensive and reasonable planning, highlighting the services for teachers, students and employing departments, and taking students' employment as the fundamental purpose.

Three functional modules, including cloud service center, teaching resource management, and remote classroom, demonstrated the implementation effect of the smart teaching system proposed in this article, completed the functional

test of the system and performed performance test on the load capacity of the database and the function of the service interface. Finally, the advantages of the system proposed in this paper are analyzed. In addition, the system services need to be tested; here, two services are selected for stress testing, resource management, and remote classroom; the specific test method is to select the two types of interfaces provided by the service read and write, in a minute time, by changing the number of clients and send requests, to obtain the average of the number of outputs per minute for each service, in pages/min. Test results are shown in Figure 8.

By analyzing the data in Figure 8, the system can still run stably and meet the basic requirements of the design under the high concurrency of the service. A brief overview of the requirements analysis and framework design to be completed in the specific implementation of the intelligent teaching system is given; then, the specific development environment of the system and the relevant details of the system service implementation are introduced; then, the effect of the implementation of the intelligent teaching system proposed in this paper is demonstrated from three functional modules, such as cloud service center, teaching resource management, and remote classroom, and the functional test of the system is completed, and the load capacity of the database. Finally, the advantages of the system proposed in this paper are analyzed.

It introduces the requirements of the digital campus academic affairs management system, analyzes the performance and requirements of the university's academic affairs network management system, and analyzes the detailed requirements of each management module and usage function; introduces the application of the digital campus academic affairs management system in the process of using, such as information resource sharing, academic completion monitoring, and teaching assessment, and optimizes the design of each management module; and carries out the interface implementation

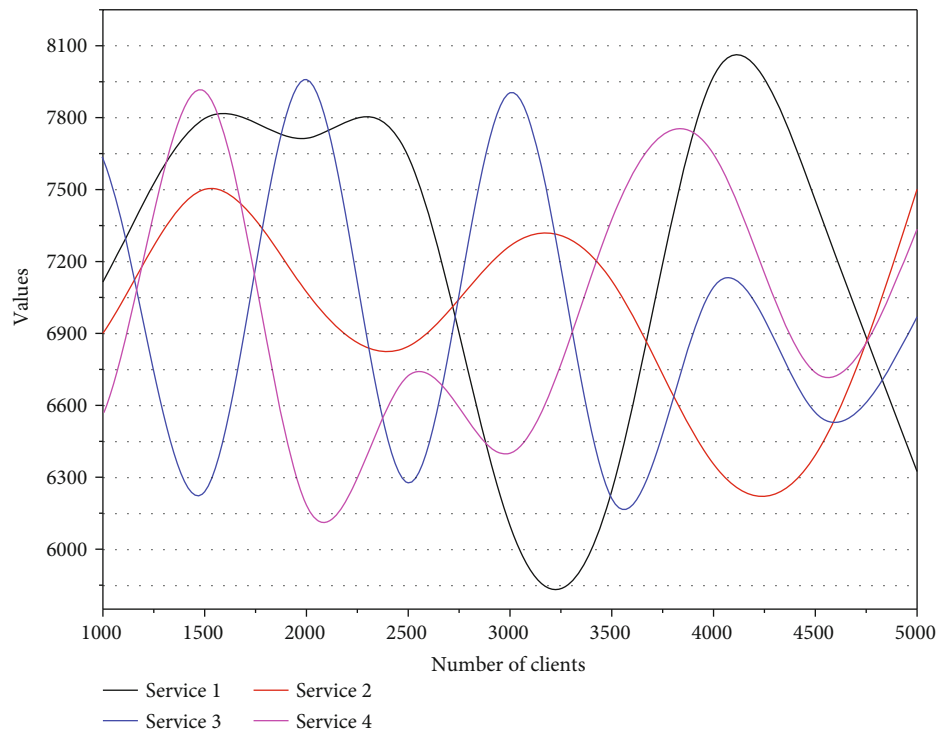


FIGURE 8: Service stress test.

and testing of various data and management rules in the process of using the academic affairs network management system developed in this project. The design of the management modules is optimized, and the interface of the academic affairs network management system developed in this project is implemented and tested in the process of using various data and management rules in the school. It is a web platform system focusing on the design of the academic affairs network management, using J2EE as the development platform and a hybrid C/S and B/S structure, and proposes an innovative information-based academic affairs network management system. Through the digital campus academic affairs management system, the sharing of information resources and application operations are realized; based on the digital campus academic affairs management system, course scheduling, learning achievement monitoring, and teachers' teaching quality assessment are carried out; it has the significance of promoting and reforming the use and development of the university academic affairs network management system.

## 5. Conclusion

This paper completes the design of the university postgraduate management information system, which meets the actual situation and requirements of postgraduate management; realizes the functions of postgraduate academic registration management, training program management, and training plan development; and provides convenience for postgraduate training management. The system started trial operation online, and the system database software was deployed on two physical servers using Oracle, and the

application server was deployed on two virtual machines on blade servers using WebSphere 6.1. The graduate student management information system adopts the concept of "big platform and small application" in the design of a digital campus system. The struts technology architecture based on J2EE architecture is designed with high efficiency and reusability, which makes the system have good maintainability and portability. While managing the system, the only data format standard belonging to the graduate system was established, which lays the foundation for the later digital campus integration construction, simplifies the steps of sharing data, and improves the operational efficiency of the system. This study focuses on the technical solution design for the needs of a single application system of the postgraduate management information system. The digital campus construction will have a large amount of personal information about the transmission, the vast majority of which can be said to carry confidential information, and the orientation of information also requires communication and coordination between school departments and authorization.

## Data Availability

The data used to support the findings of this study are included within the article.

## Conflicts of Interest




All the authors do not have any possible conflicts of interest.

## References

- [1] S. Nur Ain Basri, F. Ahmad, N. Izie Adiana Abidin et al., "Digital campus," *International Journal of Engineering & Technology*, vol. 9, no. 2, pp. 382–389, 2020.
- [2] M. N. Habib, W. Jamal, U. Khalil, and Z. Khan, "Transforming universities in interactive digital platform: case of city university of science and information technology," *Education and Information Technologies*, vol. 26, no. 1, pp. 517–541, 2021.
- [3] N. N. Minh, "The role of learning management system on university branding: evidence from Vietnam," *PalArch's Journal of Archaeology of Egypt/Egyptology*, vol. 17, no. 4, pp. 931–947, 2020.
- [4] R. G. Hadgraft and A. Kolmos, "Emerging learning environments in engineering education," *Australasian Journal of Engineering Education*, vol. 25, no. 1, pp. 3–16, 2020.
- [5] P. Paul, P. S. Aithal, and A. Bhumali, "Environmental informatics and educational opportunities in post graduate level—Indian potentialities based on international scenario," *IRA-International Journal of Management & Social Sciences*, vol. 16, no. 2, pp. 45–58, 2020.
- [6] Z. Y. Dong, Y. Zhang, C. Yip, S. Swift, and K. Beswick, "Smart campus: definition, framework, technologies, and services," *IET Smart Cities*, vol. 2, no. 1, pp. 43–54, 2020.
- [7] D. Jackson and S. Meek, "Embedding work-integrated learning into accounting education: the state of play and pathways to future implementation," *Accounting Education*, vol. 30, no. 1, pp. 63–85, 2021.
- [8] P. S. Aithal and S. Aithal, "Analysis of the Indian National Education Policy 2020 towards achieving its objectives," *International Journal of Management, Technology, and Social Sciences (IJMTS)*, vol. 5, no. 2, pp. 19–41, 2020.
- [9] A. AlShamsi, J. Mohaidat, N. al Hinai, and A. Samy, "Instructional and business continuity amid and beyond COVID-19 outbreak: a case study from the higher colleges of technology," *International Journal of Higher Education*, vol. 9, no. 6, pp. 118–135, 2020.
- [10] P. Paul, P. S. Aithal, A. Bhumali, T. Kalishankar, M. R. Saavedra, and P. S. Aremu, "Geo information systems and remote sensing: applications in environmental systems and management," *International Journal of Management, Technology, and Social Sciences (IJMTS)*, vol. 5, no. 2, pp. 11–18, 2020.
- [11] A. Tadesse, W. R. Allen, and C. Mitchell-Kernan, "Integrating educational technology in East Africa: one size does not fit all," *Monitoring of Public Opinion: Economic and Social Change*, vol. 1, pp. 91–108, 2021.
- [12] U. O. Matthew, J. S. Kazaura, and K. Haruna, "Multimedia information system (MIS) for knowledge generation and ICT policy framework in education," *International Journal of Information Communication Technologies and Human Development (IJICTHD)*, vol. 12, no. 3, pp. 28–58, 2020.
- [13] T. Anderson and P. Rivera-Vargas, "A critical look at educational technology from a distance education perspective," *Digital Education Review*, vol. 37, no. 37, pp. 208–229, 2020.
- [14] M. Hakami, "Using Nearpod as a tool to promote active learning in higher education in a BYOD learning environment," *Journal of Education and Learning*, vol. 9, no. 1, pp. 119–126, 2020.
- [15] B. A. Y. Al-Nassar, "Effect of information quality and system quality in information system success model as an antecedent of mobile learning in education institutions: case study in Jordan," *International Journal of Mobile Learning and Organisation*, vol. 14, no. 3, pp. 277–306, 2020.
- [16] F. Martin, D. Polly, S. Coles, and C. Wang, "Examining higher education faculty use of current digital technologies: importance, competence, and motivation," *International Journal of Teaching and Learning in Higher Education*, vol. 32, no. 1, pp. 73–86, 2020.
- [17] W. Wargadinata, I. Maimunah, E. Dewi, and Z. Rofiq, "Student's responses on learning in the early COVID-19 pandemic," *Tadris: Journal of Education and Teacher Training*, vol. 5, no. 1, pp. 141–153, 2020.
- [18] Y. Zuo, D. Yao, and M. Zhang, "Exploration and practice of innovation and entrepreneurship awareness embedded in experimental teaching of economic management major undergraduates: a case study from China," *Higher Education Studies*, vol. 10, no. 3, pp. 53–62, 2020.
- [19] F. Valencia-Forrester, "Models of work-integrated learning in journalism education," *Journalism Studies*, vol. 21, no. 5, pp. 697–712, 2020.
- [20] T. D. Le Hoanh Su, T. Thi-Yen-Linh, N. Thi-Duyen-Ngoc, L. Bao-Tuyen, and N. Ha-Phuong-Truc, "Development of an AI Chatbot to support admissions and career guidance for universities," *International Journal of Emerging Multidisciplinary Research*, vol. 4, no. 2, pp. 11–17, 2020.
- [21] C. A. Bonfield, M. Salter, A. Longmuir, M. Benson, and C. Adachi, "Transformation or evolution?: education 4.0, teaching and learning in the digital age," *Higher Education Pedagogies*, vol. 5, no. 1, pp. 223–246, 2020.
- [22] I. C. Utomo, S. Rokhmah, and I. Muslihah, "Web based distribution of Zakat, Infaq, and Shodaqoh (case study of Surakarta City region)," *International Journal of Computer and Information System (IJCIS)*, vol. 1, no. 1, pp. 16–21, 2020.
- [23] R. L. Quezada, C. Talbot, and K. B. Quezada-Parker, "From bricks and mortar to remote teaching: a teacher education program's response to COVID-19," *Journal of Education for Teaching*, vol. 46, no. 4, pp. 472–483, 2020.
- [24] M. Hernandez-de-Menendez, C. A. E. Díaz, and R. Morales-Menendez, "Educational experiences with generation Z," *International Journal on Interactive Design and Manufacturing (IJIDeM)*, vol. 14, no. 3, pp. 847–859, 2020.
- [25] C. Z. Levkoe, I. Knezevic, D. Appavoo, A. Moraes, and S. Scott, "Serving up food studies online: teaching about "food from somewhere" from nowhere," *Food, Culture & Society*, vol. 23, no. 3, pp. 434–453, 2020.
- [26] M. Hamadi, J. el-den, C. Narumon Sriratanaviriyakul, and S. Azam, "A social media adoption framework as pedagogical instruments in higher education classrooms," *E-Learning and Digital Media*, vol. 18, no. 1, pp. 55–85, 2021.
- [27] L. Yu, S. Tao, W. Gao, and L. Yu, "Self-monitoring method for improving health-related quality of life: data acquisition, monitoring, and analysis of vital signs and diet," *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 1, pp. 24–31, 2021.

## Research Article

# An Improved MOEA/D Algorithm for Complex Data Analysis

Weihoa Qian <sup>1</sup>, Jiahui Liu,<sup>1</sup> Yuanguo Lin,<sup>1</sup> Lvqing Yang <sup>1</sup>, Jianwei Zhang <sup>2</sup>, Hang Xu,<sup>3</sup>  
Minghong Liao,<sup>1</sup> Yuxuan Chen,<sup>1</sup> Yunyi Chen,<sup>1</sup> and BingShuai Liu<sup>1</sup>

<sup>1</sup>Xiamen University, China

<sup>2</sup>Zhengzhou University, China

<sup>3</sup>Putian University, China

Correspondence should be addressed to Lvqing Yang; [lqyang@xmu.edu.cn](mailto:lqyang@xmu.edu.cn) and Jianwei Zhang; [mailzjw@163.com](mailto:mailzjw@163.com)

Received 14 July 2021; Revised 30 August 2021; Accepted 1 September 2021; Published 29 September 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Weihoa Qian et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

There are a large number of multiple level datasets in the Industry 4.0 era. Thus, it is necessary to utilize artificial intelligence technology for the complex data analysis. In fact, the technology often suffers from the self-optimization issue of multiple level datasets, which is taken as a kind of multiobjective optimization problem (MOP). Naturally, the MOP can be solved by the multiobjective evolutionary algorithm based on decomposition (MOEA/D). However, most existing MOEA/D algorithms usually fail to adapt neighborhood for the offspring generation, since these algorithms have shortcomings in both global search and adaptive control. To address this issue, we propose a MOEA/D with adaptive exploration and exploitation, termed MOEA/D-AEE, which adopts random numbers with a uniform distribution to explore the objective space and introduces a joint exploitation coefficient between parents to generate better offspring. By dynamic exploration and joint exploitation, MOEA/D-AEE improves both global search ability and diversity of the algorithm. Experimental results on benchmark data sets demonstrate that our proposed approach achieves global search ability and diversity in terms of the population distribution than state-of-the-art MOEA/D algorithms.

## 1. Introduction

A multiobjective optimization problem (MOP) refers to the determination of a vector composed of decision variables in a feasible domain, which satisfies all constraints and optimizes a vector composed of multiple objective functions. The multiple objective functions of these component vectors usually contain subobjectives that conflict with each other. The improvement of one subobjective may lead to the performance degradation of another or several subobjectives. That is, it is impossible to achieve the optimal values for multiple subobjectives; rather, there must be coordination and compromise among them so that each subobjective can be optimized as well as possible. MOPs are a common kind of optimization problems in the real world and often mentioned in the field of scientific research and its applications.

Many scholars have tried to solve such problems by using evolutionary algorithms. The first step is to simplify the problem of two dimensions into a single-objective prob-

lem by setting weights. However, due to the increase in the dimensions of the objective function, the simple weighting method cannot reflect the real situation of the population, and the distribution of the population is uneven. Then, scholars continued to study the strategy of using multiobjective evolutionary algorithms (MOEAs) and proposed the classic nondominated genetic sorting algorithm II (NSGA-II) [1]. This algorithm is based on the method of fast nondominated set sorting and congestion calculation. It has good performance for low-dimensional multiobjective optimization problems, but the congestion distance is not suitable for high-dimensional space, and the computational complexity is relatively high. To speed up the convergence of the algorithm, scholars proposed the classic multiobjective evolutionary algorithm based on decomposition (MOEA/D) [2] in 2007. The decomposition-based method improves the convergence speed of the algorithm, which reduces the computational complexity, and obtains more uniform solutions than NSGA-II. Thus, MOEA/D has been widely used in

complex data analysis [3, 4]. In recent years, with the development of Industry 4.0, many combinatorial optimization problems about large-scale and multiple level datasets have arisen [5, 6], while the MOEA/D algorithm has been introduced to address this problems, which is a kind of complex data analysis MOPs. For example, in the MOEA/D-Lévy algorithm [7], a Lévy flight is a short-distance hopping exploratory search strategy with occasional long-distance development search. It may be an ideal search strategy [8], since it can improve population diversity to prevent premature convergence and jump out of local optimal solutions [9]. However, the frequency of generating better offspring in the later iterations of the algorithm is very low. Besides, MOEA/D algorithms usually suffer from adapting neighborhood for the offspring generation, since the global search ability and the parent adaptive ability are insufficient.

To address the shortcomings of the above algorithms, we propose a MOEA/D with Adaptive Exploration and Exploitation, named MOEA/D-AEE, which uses dynamic search and joint development to solve global search and adaptive problems, and thereby achieves better sparse distribution and global search ability. Specifically, the proposed algorithm uses uniformly distributed random numbers to explore the solution space and introduces the joint exploitation coefficient between parents to produce better offspring by adaptively adjusting parent relationships. When two selected parents are similar, it adopts an active exploration strategy to select the parent from the initial population. In contrast, when the two selected parents are quite different, the algorithm gradually employs the exploitation strategy to select the parent from the neighbor. To verify the performance of the proposed algorithm, two experimental schemes are proposed in this study, and a comparative test is carried out on the three OR-Library public data sets in PO. The traditional MOEA/D algorithm, evolution algorithm, NSGA-II algorithm, and Lévy flight mutation algorithm and the global search performance of the algorithms are compared. The experimental results and analysis show that the performance of our algorithm is advanced in the six metrics, and the global search ability is further improved, especially in the later stage of iteration. There is still a long trajectory update, the population has a faster convergence speed in the first 20 generations and converges to the Pareto front, and the diversity performance is better on the three test platforms.

The remainder of this paper is organized as follows. Section 2 briefly reviews the research progress. In Section 3, we describe the method and implementation of the algorithm. Section 4 provides the experimental results, and the corresponding discussions are given in Section 5. Section 6 summarizes this paper and proposes possible future directions.

## 2. Related Work

**2.1. MOEA/D Algorithms.** MOPs are a class of optimization problems that contain multiple conflicting objectives. Compared with the single-objective optimization problem (SOP), the goal of an MOP is to obtain a set of optimal solutions showing the best trade-off between its multiple objectives

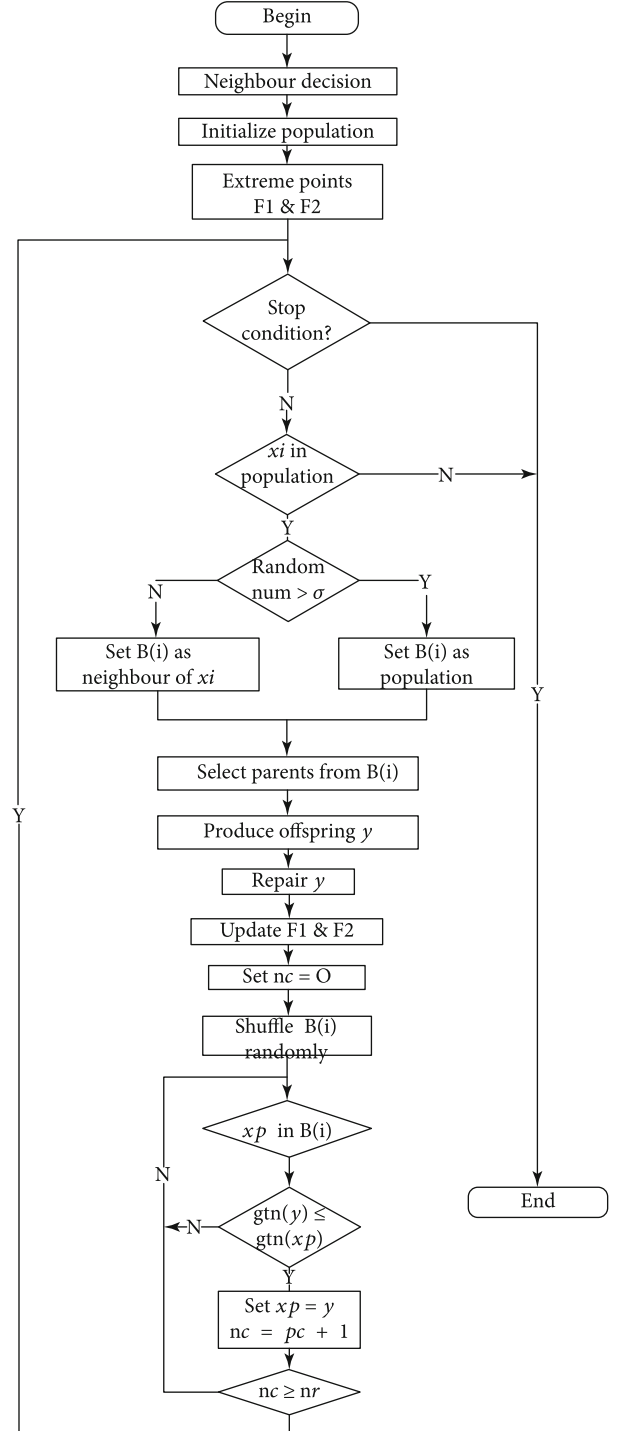


FIGURE 1: MOEA/D-AEE algorithm flow chart.

rather than an optimal solution. Such an optimal-set Pareto front contains all feasible solutions in the feasible problem domain that are not governed by other solutions. The MOP solution algorithm is called MOEA; Deb et al. and others proposed the classical NSGA-II [1], and this kind of problem was solved by sorting noncollective sets. Then, Zhang and Li proposed an MOEA/D, decomposing multiobjective problems into subproblems of multiple single-objective problems and optimizing them at the same time.



```

1 Select Neighbour B;
2 Initialize population P;
3 Compute extreme points F1 and F2;
4 while stop condition not reached do
5   for xi in opulation do
6     if random num < σ then
7       Set B(i) as neighbor of xi;
8     else.
9       Set Set B(i) as population;
10    end.
11    Select parents from B(i).
12    Use proposed method to reproduce offspring y;
13    Repair y to reasonable range;
14    Update extreme points F1 and F3;
15    Set update counter nc = 0;
16    Shuffle B(i) randomly;
17    for xp in neighbor of B(i) do
18      if gtn(y|λ, rp) < = gtn(xp|λ, rp) then
19        Set xp = y;
20        Set nc + 1;
21        if nc > = nr then
22          Break;
23        end
24      end
25    end
26  end
27 end

```

ALGORITHM 1: MOEA/D-AEE pseudocode description. MOEA/D-AEE.

Common decomposition methods in this method are the weighted sum method, Chebyshev method, and boundary crossing method [2]. The formula of the most common Chebyshev decomposition method is as follows:

$$\min_x g^{te}(x|\lambda, z^*) = \max_{m=1, \dots, M} \{\lambda_m |f_m(x) - z_m^*|\}, \quad (1)$$

$$\lambda = (\lambda_1, \dots, \lambda_M), \lambda_1 + \dots + \lambda_M = 1, \quad (2)$$

$$z^* = (\min f_1, \dots, \min f_M). \quad (3)$$

$z_m^*$  in the objective space of  $m$  dimensions represents the minimum value on the  $m$  objective dimensions,  $\lambda$  represents the weight vector, and  $z$  represents the reference point. The intersection of each dimension contour is best in the direction of the weight vector, and these points ultimately form the Pareto front.

To solve the problem of the inability to address a disproportionate scaling objective with the ordinary Chebyshev method, this paper adopts the narrowband interference-(NBI-) Chebyshev decomposition method [10] proposed by Zhang et al. The mathematical formula is as follows:

$$r^{(i)} = \alpha_i F^1 + (1 - \alpha_i) F^2, \alpha_i = \frac{N - i}{N - 1}, \quad (4)$$

$$g^{tn}(x|r^{(i)}, \lambda) = \max \left\{ \lambda_1 (f_1(x) - r_1^{(i)}), \lambda_2 (f_2(x) - r_2^{(i)}) \right\}, \quad (5)$$

$$\lambda_1 = |F_2^2 - F_2^1|, \lambda_2 = |F_1^2 - F_1^1|, \quad (6)$$

where  $F^1$  and  $F^2$  are the two extreme points of the Pareto front in the objective space, and there are  $n$  reference points  $r^i$ ,  $i = 1, \dots, N$ , which represent  $n$  reference points arranged on a line between the two extreme points.  $\lambda_1$  and  $\lambda_2$  are the weight vectors calculated by  $F^1$  and  $F^2$ , respectively, which are perpendicular to the line between the two extreme points.

In recent years, scholars have made many improvements to MOEA/D.

First of all, some of them have made improvements to the algorithm itself. For the drawback of the allocation of resources, Zhang et al. proposed an algorithm to dynamically allocate computational resources to different subproblems in MOEA/D called the MOEA/D-DRA algorithm [11]. Each subproblem uses an efficiency function to measure the improvement rate of the aggregation function and only calculates a subproblem with a higher efficiency function each time. In order to deal with the complexity of the Pareto set, Li and Zhang presented a new version of MOEA/D based on differential evolution, known as the MOEA/D-DE algorithm [12]. The test showed that MOEA/D was significantly better than NSGA-II, indicating that decomposition-based multiobjective evolutionary algorithms are very promising for addressing complex Pareto set PS shapes. To improve selection matching process, Li et al. proposed a simple and effective stable matching model (STM) [13]. Each subproblem was matched with a solution

TABLE 1: Specific parameter configuration for the algorithm.

Algorithm	Parameters	Parameters	Parameters
MOEA/D-AEE	$\alpha_0 = 1e - 05$	$\beta = 0.3$	Mutation rate = $1/n$
MOEA/D-lévy	$\alpha_0 = 1e - 05$	$\beta = 0.3$	Mutation rate = $1/n$
MOEA/D-GA	Crossover rate = 0.7	Mutation rate = 0.01	
NSGA-II	Crossover rate = 0.7	Mutation rate = 0.01	
MOEA/D-DEM	$F = 1.3$	Mutation rate = $1/n$	
MOEA/D-DE	$F = 1.3$		

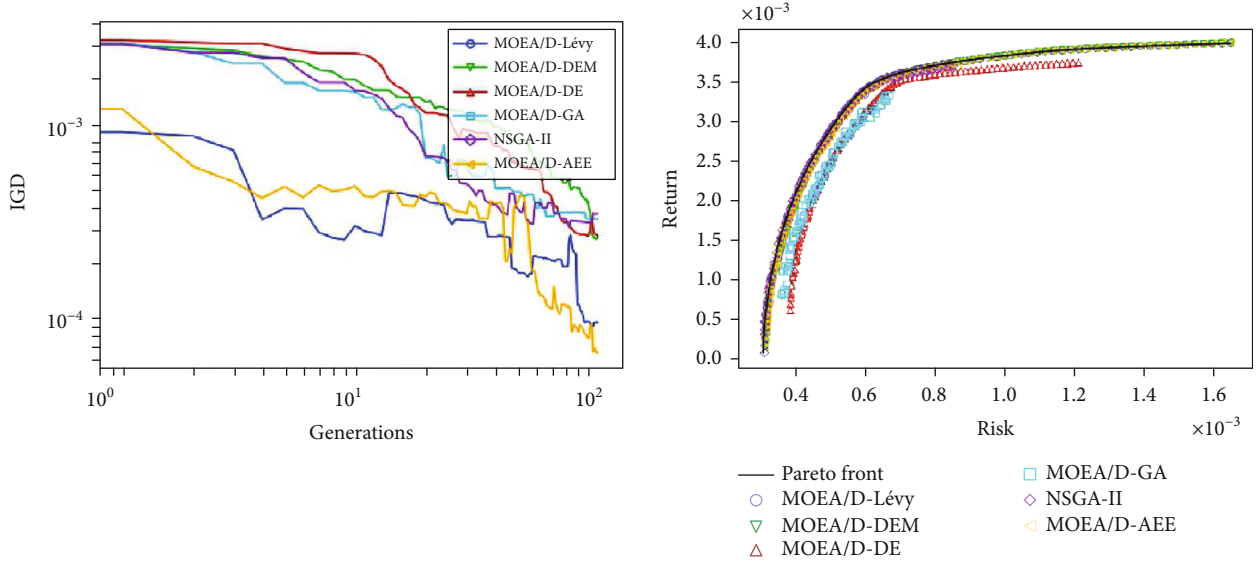


FIGURE 2: Nikkei IGD indicators and population distribution.

using an STM model, and the convergence and diversity of the algorithm were balanced.

Secondly, for the adaptive problems, Wang et al. proposed an adaptive evolutionary strategy AES-MOEA/D algorithm [14] based on an adaptive evolutionary strategy, and three improved strategies were used: the evolution strategy of simulated binary crossover (SBX) competition with the DE operator, the adaptive adjustment strategy for the mutation probability, and the boundary optimization strategy of double-sided mirror theory, improving diversity in late evolution can improve the performance of MOEA/D algorithms. Chiang and Lai also considered adaptive ability, and they proposed an algorithm based on the MOEA/D adaptive matching selection mechanism called the MOEA/D-AMS algorithm [15]. By distinguishing between resolved and unresolved subproblems, they select only unresolved subproblems to match each time. The matching pool size is dynamically adjusted according to the distance in the objective space.

Moreover, for the purpose of expanding the application of MOEA/D in high-dimensional objectives, Zheng et al. proposed an improved MOEA/D algorithm called the I-MOEA/D algorithm [16]. Compared to the original MOEA/D, the adjusted weight vector makes the distribution

of solutions more extensive and more effective in ensuring the diversity and convergence of solutions in the objective space. Tan et al. also improved the MOEA/D algorithm in high dimensions and proposed a MOEA/D-based uniform design algorithm called the UMOEA/D algorithm [17]. The weight vectors on three to five objective spaces are more uniform.

Besides, in some existing MOEAs, an external archive is usually adopted to save elite solutions found during the evolutionary search to help improve the population diversity. Inspired by the decomposition strategy, Zhang et al. propose an efficient archiving approach, called the decomposition based archiving approach (DAA) [18] to prune the external archive size. In DAA, the size of the archive depends on the number of subspaces, and they do not need to compare new solutions with all the archived solutions. This can reduce the computational cost significantly. For the disadvantages of single archive, they also study a multiobjective feature selection approach, called two-archive multiobjective artificial bee colony algorithm (TMABC-FS) [19] to balance the distribution and convergence of solutions. The leader archive and the external archive are employed to enhance the search capability of different kinds of bees to avoid miss potentially important regions.

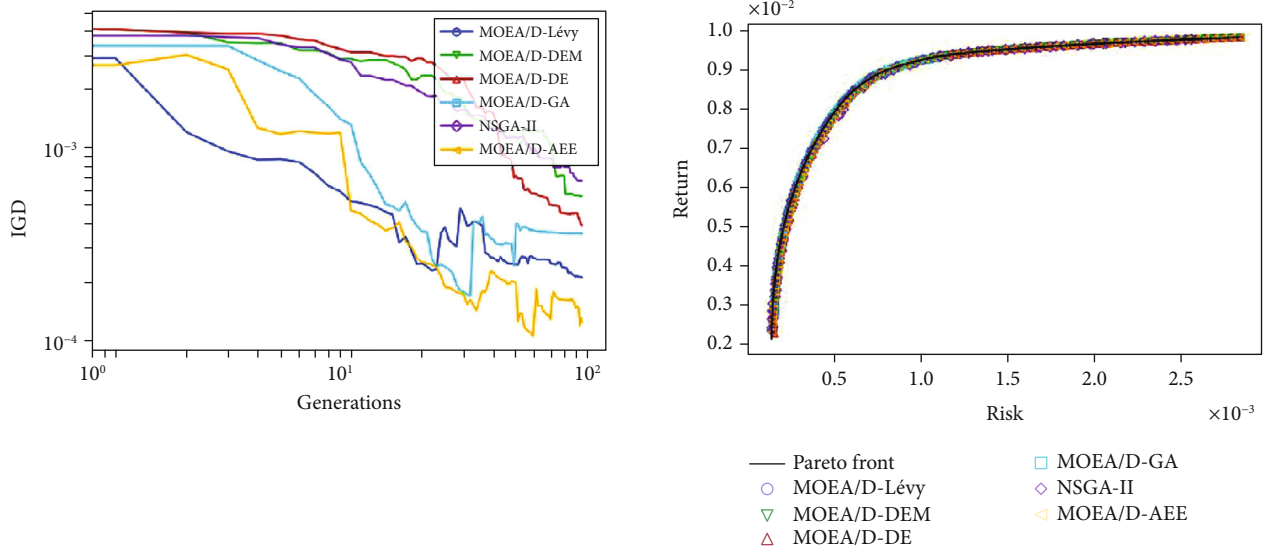


FIGURE 3: DAX 100IGD indicators and population distribution.

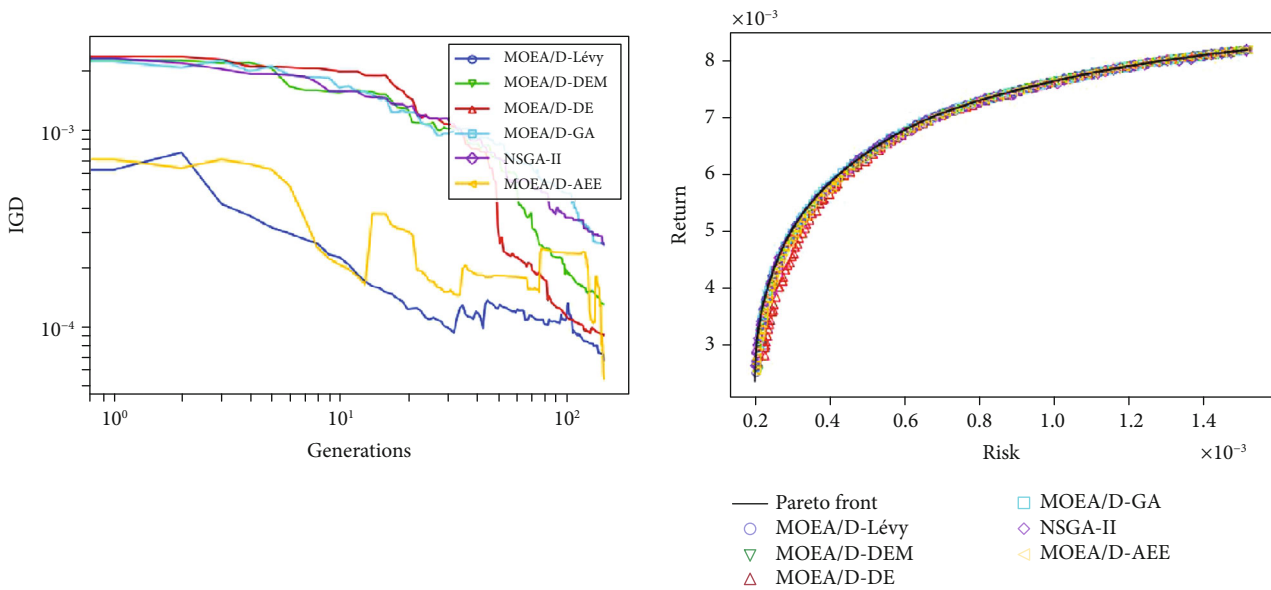


FIGURE 4: FTSE 100IGD indicators and population distribution.

2.2. MOEA/D in Industry 4.0. The self-optimization problem of cyber physical systems (CPSs) can be considered as a kind of MOPs; therefore, the MOEA/D algorithm is the need for automated and enhanced decision makings.

On the one hand, metaheuristics have been used successfully to address realistic planning and scheduling problems. Cota et al. propose multiobjective extensions of the adaptive large neighborhood search (ALNS) metaheuristic with Learning Automata (LA) to improve the search process and to solve the large scale instances efficiently. Two new algorithms are MO-ALNS and the MO-ALNS/D, and they address the unrelated parallel machine scheduling problem to maximize production [4]. Energy efficiency and timeliness

are two mutually contradictory objectives, and Shukla et al. propose a novel approach, based on the popular multiobjective evolutionary algorithm, Nondominated sorting genetic algorithm-II, to solve this problem. They introduce novel algorithms for membership function generation and calculation of fuzzy earliness [20]. Fu et al. study an integrated scheduling problem in a distributed manufacturing system [21]. They build a stochastic multiobjective model with consideration of two criteria of scheduling and uncertainties of the operation process. Also, they design a multiobjective brain storm optimization algorithm to solve the model. Fu et al. also investigate a flow-shop scheduling problem under the consideration of multiple objectives, time-dependent

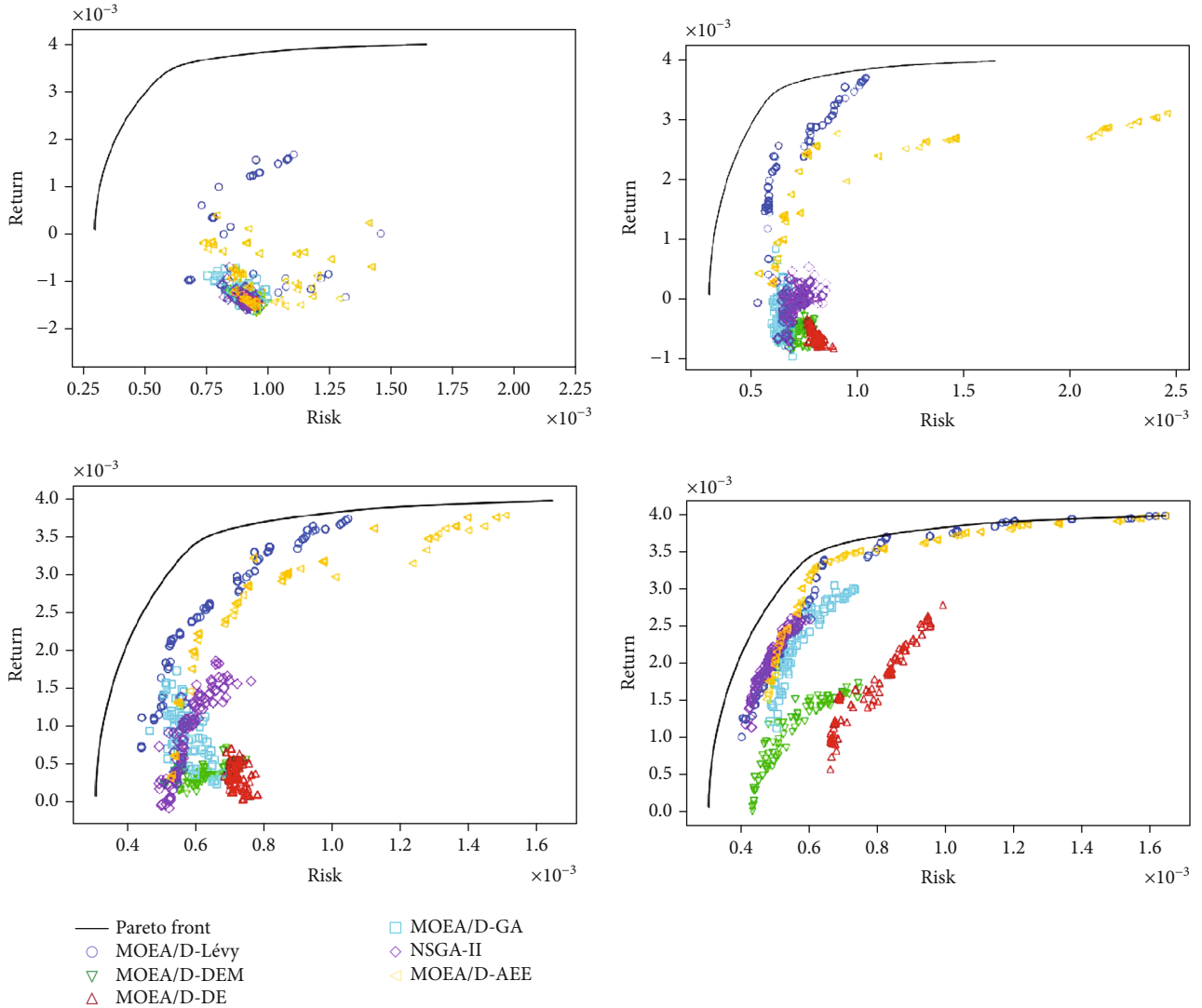


FIGURE 5: Distribution of 1, 10, 20, and 50 generations in the experiment.

processing time and uncertainty. A mixed integer programming model is formulated for this problem, and a fireworks algorithm is developed [22].

On the other hand, maintenance has always been a key activity in the manufacturing industry, and Goti et al. focus on condition-based maintenance (CBM) optimization in an industrial environment, with the objective of determining the optimal values of preventive intervention limits for equipment under corrective and preventive maintenance cost criteria. They develop a cost-benefit mathematical model. It considers the evolution in quality and production speed, along with condition based, corrective, and preventive maintenance [23]. Besides, existing big service selection and composition algorithms in industry fail to achieve the global optimum solution in a reasonable time, and Dutta et al. design an efficient quality of service-aware big service composition methodology using a distributed coevolutionary algorithm. They develop a distributed NSGA-III for finding the optimal Pareto front and a distributed multiobjective Jaya algorithm for enhancing the diversity of solutions [24].

Rather than other MOEA/D algorithms such as [25, 26], our algorithm employs both global exploration and local exploitation and thereby solves the self-optimization issue of multiple level datasets in the Industry 4.0 era. We introduce the MOEA/D-AEE algorithm in the following section.

### 3. Methods

**3.1. Framework of MOEA/D-AEE.** In this section, we describe the MOEA/D-AEE algorithm, which integrates Lévy flight algorithms with dynamic search and joint development capabilities in MOEA/D multiobjective optimization algorithms. The complete diagram of the algorithm is shown in Figure 1. In this paper, a vector is used to represent the problem variables in investment optimization. The solution of the algorithm is used with a uniform distribution in  $(0, 1)$  for initialization. For the generated progeny, this paper uses a new normal boundary intersection- (NBI-) Chebyshev decomposition method in the MOEA/D decomposition to address the return and risk [10] of the two objective

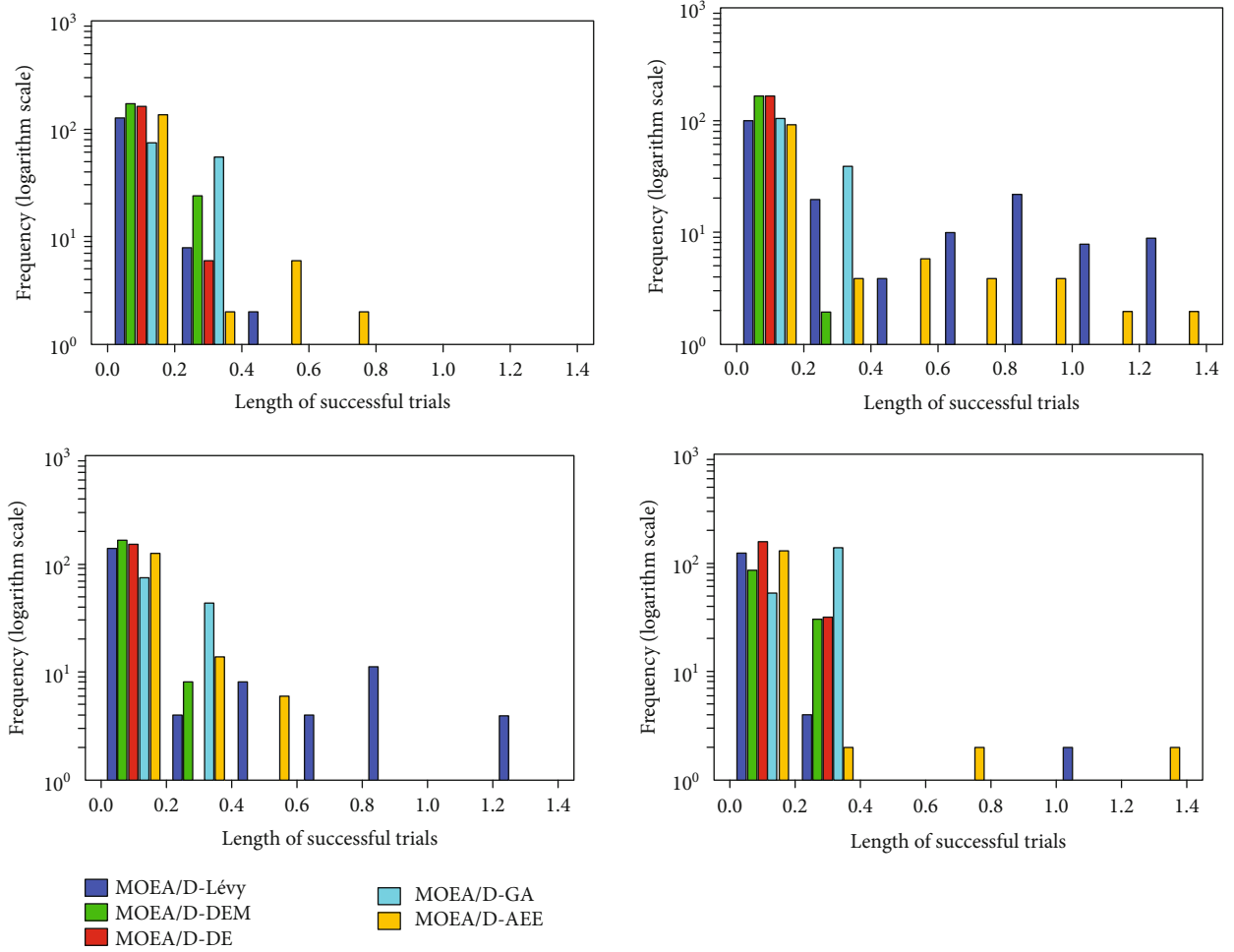


FIGURE 6: Diagram of the 1-, 3-, 5-, and 10-generation successful update trajectories.

functions. At the same time, this paper adopts the diversity maintenance strategy, the iterative mutation method of the Lévy flight, and polynomial mutation based on joint coefficients.

The Lévy flight mutation operator used in this paper is similar to the DE mutation operator and is used on different individuals. The difference is that the scaling factor used for Lévy flight mutation is generated by a heavy-tailed distribution rather than a fixed value. Additionally, Lévy flight mutation usually requires only two parents, not three. The Lévy flight mutation formula is as follows:

$$\text{Levy}(\beta) \sim \frac{u}{|v|^{1/\beta}}, \quad (7)$$

$$u \sim N(0, \sigma_u^2), v \sim N(0, \sigma_v^2), \quad (8)$$

$$\sigma_u = \left[ \frac{\Gamma(1+\beta) \sin(\pi\beta/2)}{\Gamma(1+\beta/2)\beta 2^{\beta-1/2}} \right]^{1/\beta}, \sigma_v = 1, \quad (9)$$

where  $\beta$  is the parameter of the Lévy flight mutation algorithm,  $u$  is a random variable with a mean value of 0 and variance of  $\sigma_u^2$ , and  $V$  is a random variable with a mean

value of 0 and variance of  $\sigma_v^2$ . The expressions of  $\sigma_u$  and  $\sigma_v$  are shown in formula (9), where  $\Gamma$  is the gamma function.

The Lévy flight algorithm, because of its lack of global search ability and adaptive control ability, has a low frequency of generating better offspring in the later stage of evolution, thus reducing the effectiveness of portfolio optimization. To address these issues, we introduce a dynamic coefficient to explore the objective space globally, which improves the objective search ability. On the other hand, we utilize the joint coefficient to adaptively control exploration and exploitation for parent selection, which leads to the convergence of the algorithm.

**3.2. Global Search and Adaptive Control Ability.** The new solution generated in the subproblem in the MOEA/D algorithm replaces the old solution only in the corresponding neighborhood  $B(i)$ . The resulting new solution may not be suitable for the subproblem in the neighbourhood; so, the algorithm can easily fall into a local optimum. The objective search ability is very important for the exploration of the solution space. A strong objective search ability can make the algorithm jump out of a local optimal solution and find a better global optimal solution. We argue that the dynamic diversity of the algorithm is helpful in improving the

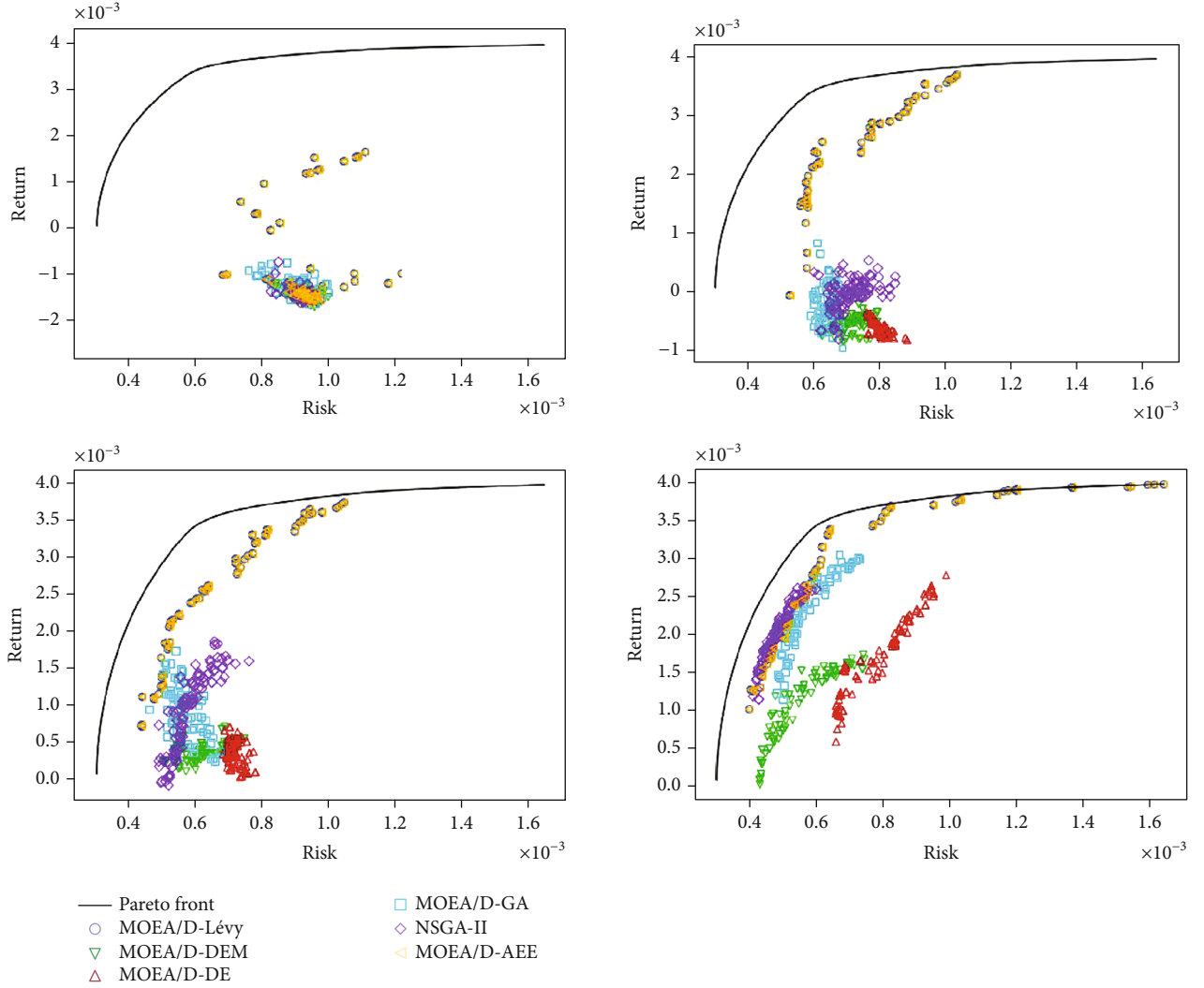


FIGURE 7: Population distribution (epsilon = 0.5).

objective search ability. To this end, we adopt a dynamic stochastic method to explore the objective space globally. That is, the amount of exploration is controlled globally by a dynamic parameter,  $\varepsilon(i)$ , that determines the randomness in parent selection. Thus, this objective search ability brings the offspring closer to the Pareto front. The mutation formula is composed of a nested polynomial mutation of the improved Lévy flight algorithm, and the formula is shown in (10), where  $x^i$  and  $x^j$  are the two parents selected in this study,  $y$  is the resulting offspring,  $\oplus$  is the method of multiplying item by item,  $\alpha_0$  is the scaling factor used, and Lévy ( $\beta$ ) is the vector generated by the MA (Mantegna's algorithm). The numbers are generated randomly under a symmetrical heavy-tailed distribution, and the algorithm data will be described in the fourth part of experiment 1. Our proposed method improves the global search ability by automatically adapting a dependent exploration probability,  $\varepsilon(i)$  at the beginning to obtain a global parameter by hand. The formula is as follows:

$$y = x^i \varepsilon + \alpha_0 (x^i - x^j) \oplus \text{Levy}(\beta). \quad (10)$$

Although the global search ability mentioned above can prevent the algorithm from falling into a local optimum, it cannot guarantee the convergence of the algorithm because it is easy to cause the algorithm to fail to converge by continuously exploring the solution space. This involves a trade-off between exploration and exploitation. The aim is to design an adaptive algorithm that can not only explore the generation of parents that are beneficial to the global optimal solution but also make full use of the current better generated parent.

An ideal solution is to automatically develop the solution space between the two parents according to the solution space under the constraint of the joint coefficient. Inspired by the value difference-based exploration (VDBE) method [27], we randomly adjust the parent relationship adaptively from uniformly distributed random variables. Specifically,

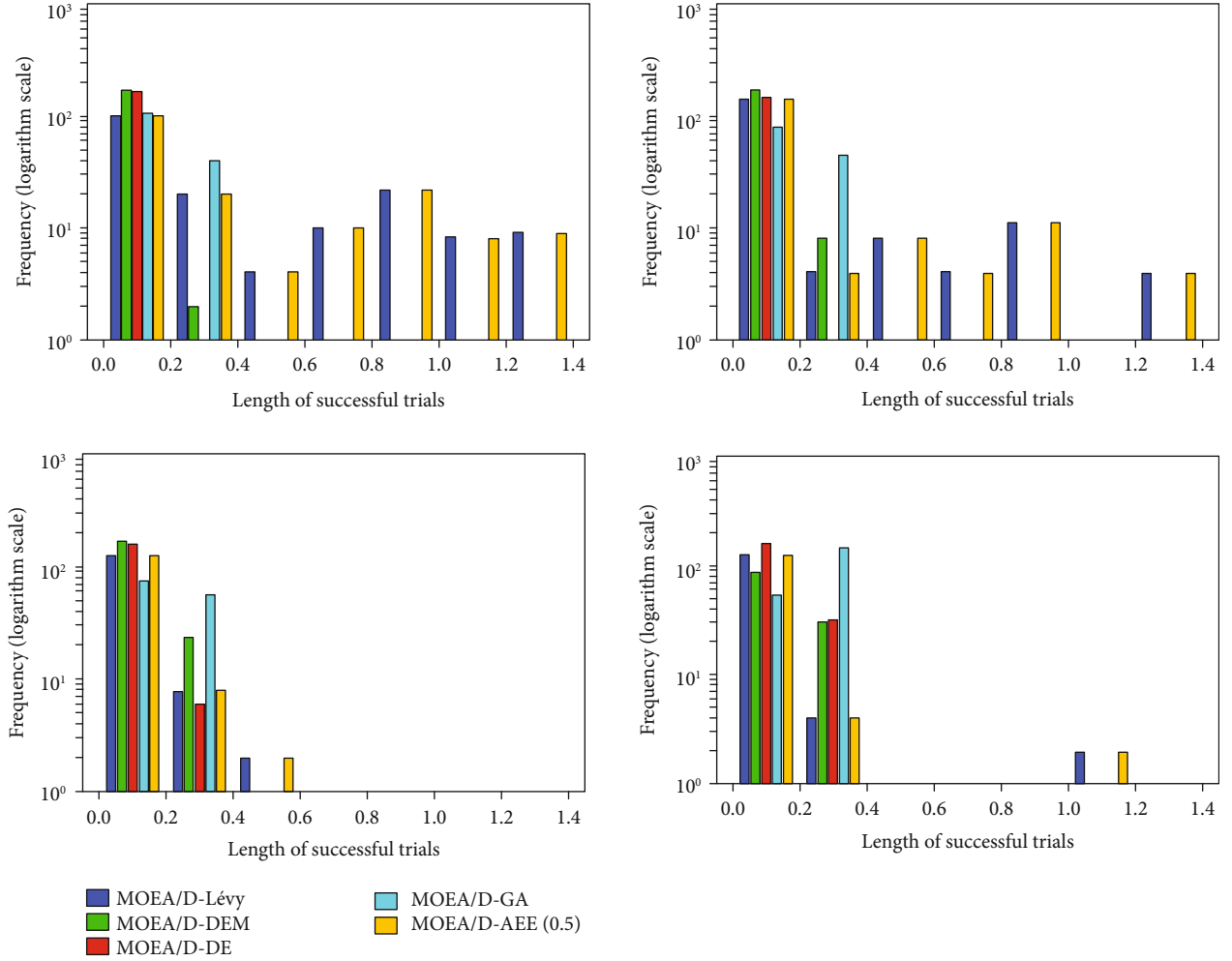


FIGURE 8: Successfully updated trajectory diagram (epsilon = 0.5).

we choose a random number epsilon between 0 and 1, and each time, we randomly select one of the parents with a probability of epsilon. If epsilon is less than  $\sigma$ , we select the parent from the neighbor; otherwise, we select the parent from the initial population. For choosing another parent in the initial population with a probability of (1-epsilon), the formula is as follows:

$$y = x^i \varepsilon + \alpha_0 (x^i - x^j) (1 - \varepsilon) \oplus \text{Levy}(\beta). \quad (11)$$

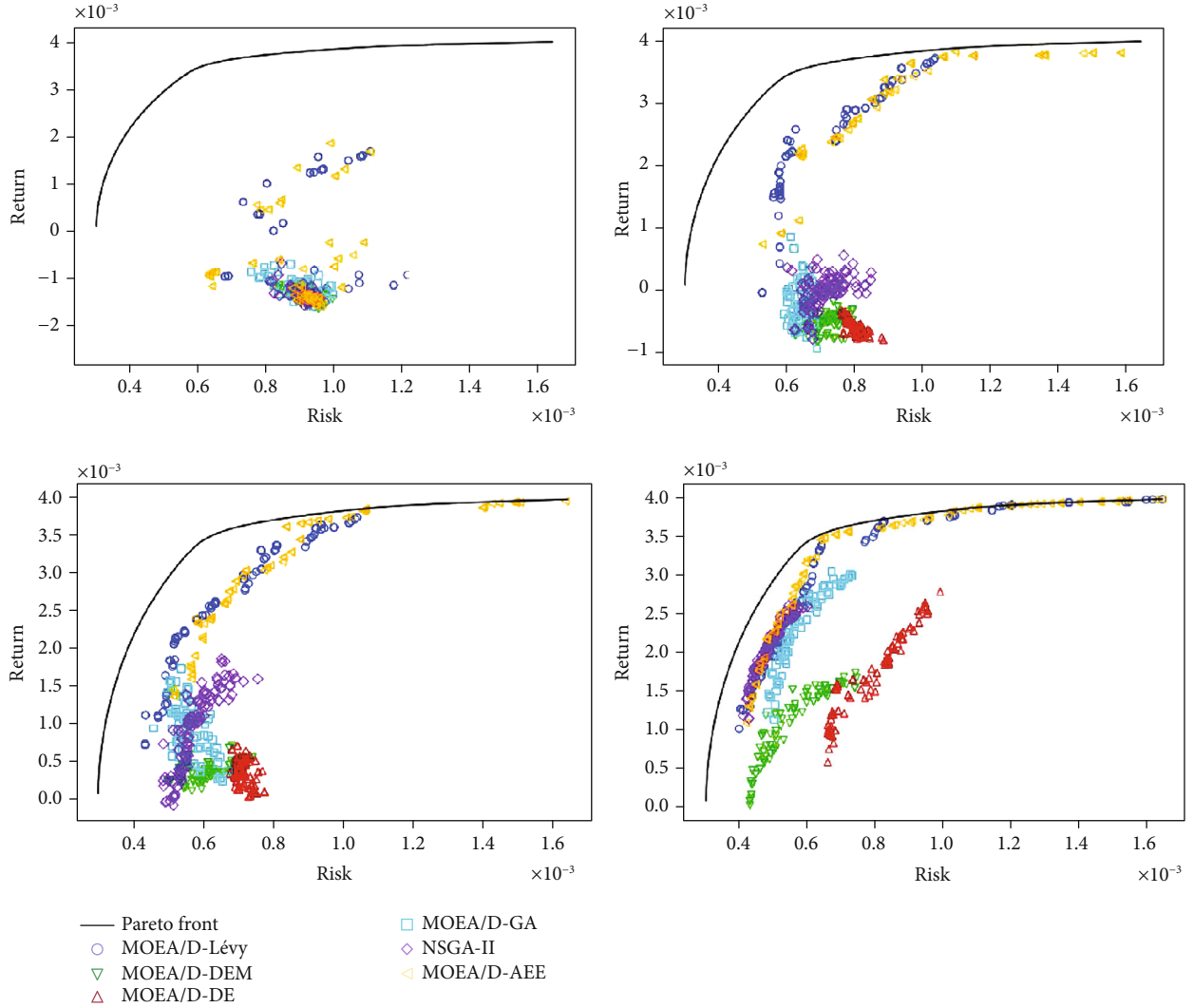
Our proposed method can achieve adaptive control of the degree of preference for exploration and exploitation with epsilon since each decision of whether to generate an offspring is explored by  $\varepsilon(i)$  availability and expanded by  $(1 - \varepsilon(i))$  availability. When two selected parents are similar, it adopts an active exploration strategy to select the parent from the initial population. In contrast, when the two selected parents are quite different, the algorithm gradually employs the exploitation strategy to select the parent from the neighbor. In this way, the algorithm can simultaneously search globally and converge. To prove the effectiveness of the algorithm, we select several sets of fixed joint coefficients

to develop and compare with the proposed method. The details are described in the fourth part of experiment 2.

**3.3. Algorithm Implementation.** This section embeds the above AEE method into the MOEA/D algorithm to form the MOEA/D-AEE algorithm. The pseudocode is shown in Algorithm 1 below.  $B$  represents the generated neighbors,  $P$  represents the generated initial population,  $F^1$  and  $F^2$  are the two extreme points of the Pareto front in the objective space, and  $\sigma$  adopts the diversity maintenance strategy in [2] by selecting the generated parent from the neighbors when it is less than  $\sigma$  and otherwise selecting the generated parent from the population. For the reproduction operation, this study uses the optimized Lévy algorithm nested polynomial mutation to generate offspring; the  $g^m$  Chebyshev discriminant method representing the NBI type introduced in part 2 is used, where  $r^p$  and  $\lambda$  can be calculated by formula (6), and  $n_r$  updates the upper bound of the neighbors.

Here are the detailed steps of the MOEA/D-AEE algorithm:

*Step 1.* Select the neighbor  $B$  based on parameters.

FIGURE 9: Population distribution ( $\epsilon = 0.2$ ).

*Step 2.* Initialize the population  $P$  based on parameters and compute objective [return, risk].

*Step 3.* Initialize the extreme points of the Pareto Front  $F^1$  and  $F^2$  based on objective [return, risk].

*Step 4.* Start the iteration of the algorithm based on parameters.

*Step 5.* Each iteration traverses each individual in the population to start execute.

*Step 6.* A random number will be generated and compared with the threshold  $\sigma$ . If the random number  $> \sigma$ , the entire population will be used as the mating pool; otherwise, the neighbors of the individual will be selected as the mating pool.

*Step 7.* According to the formula (11), use the algorithm proposed in this paper to generate the offspring  $y$  on line 12 in the pseudocode and obtain the objective [return, risk] of  $y$ .

*Step 8.* Update the extreme points of the Pareto Front  $F^1$  and  $F^2$ .

*Step 9.* Set counter  $nc$  and shuffle mating pool randomly.

*Step 10.* According to formula (5), compute the  $gtn$  value to update the individual and objective [return, risk].

*Step 11.* Exit the loop and return the updated objective [return, risk] when reach the criterion of convergence or reach the upper limit of the counter.

In addition, the offspring generated by the MOEA/D-AEE algorithm is likely to be repaired because the reproduction operation may not satisfy the constraint. Therefore, the steps of the repair method can be defined as follows:



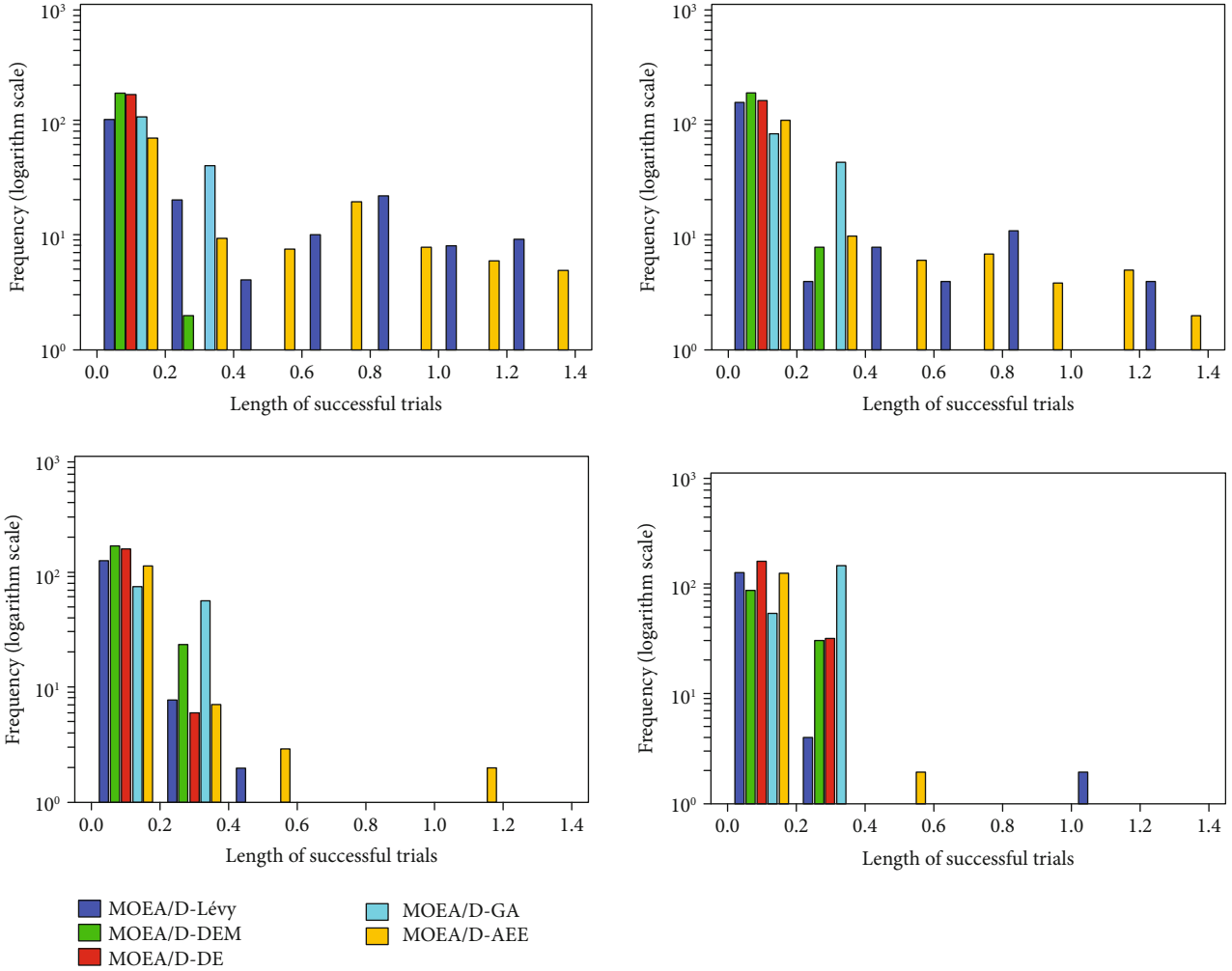


FIGURE 10: Successfully updated trajectory diagram (epsilon = 0.2).

*Step 1.* Traverse all the offspring, set all the negative offspring to 0, and keep the other offspring unchanged.

*Step 2.* Calculate the sum  $s$  of all offspring vectors.

*Step 3.* Traverse all the offspring again and scale each offspring vector; that is,  $y_i = y_i/s$ , so that the summation of all vectors equals to 1.

The algorithm we proposed needs to traverse each individual of the population in each iteration, so that the time complexity should be  $O(I \times P)$ , where  $I$  represents the number of iterations of the population, and  $P$  represents the size of the population.

It should be noted that the parameter settings of our algorithm and other comparison algorithms in the experiment 1 in chapter 4 are the same; that is, the number of iterations is 1500, and the population size is 100. It can be seen that the complexity of this algorithm is comparable to other algorithms, and it has not become significantly more complicated.

## 4. Experiments

*4.1. Preparation and Setup of Experiments.* The experiments are arranged in two parts, mainly to analyze the performance of the MOEA/D-AEE algorithm proposed in this study, and two experiments are carried out, as shown in 4.2 and 4.3. Experiment one compares several mainstream MOEA/D-based evolution algorithms, NSGA-II algorithms, Lévy flight mutation algorithms, and MOEA/D-AEE algorithms, while experiment two compares the effects of dynamic and static joint coefficients on the adaptability of MOEA/D-AEE algorithms.

*4.1.1. Data Sets.* One frequently used benchmark in assessments is the OR-library, which contains five PO data sets: Hangseng, DAX 100, FTSE 100, S&P 100, and Nikkei. The experiment was performed on three data sets disclosed by the OR-library, which contain 225, 85, and 89 test samples; the corresponding FTSE 100 data are from Japan Nikkei, Germany, and the United Kingdom; data set source: <http://people.brunel.ac.uk/~mastjjb/jeb/orlib/portinfo.html>.

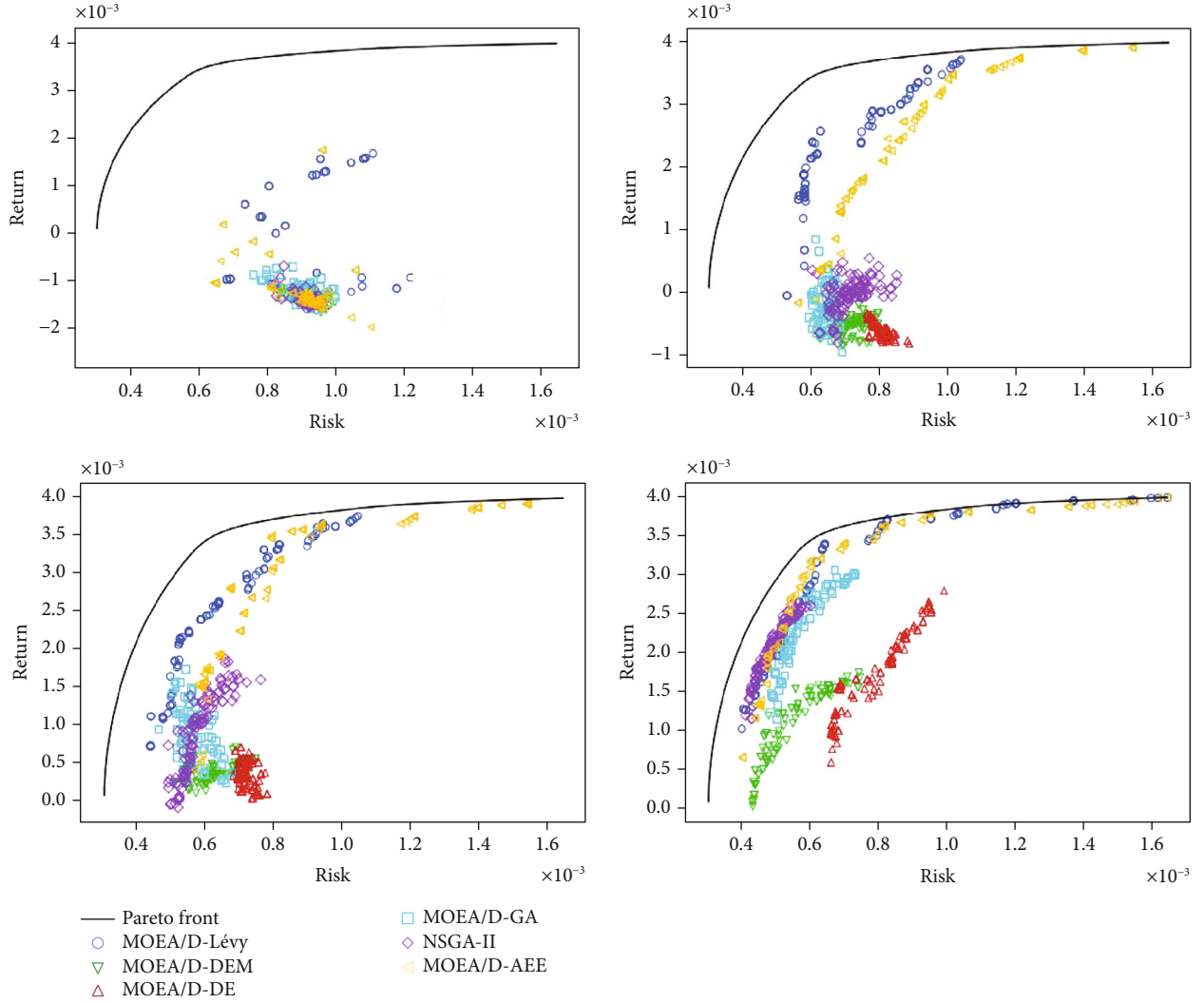


FIGURE 11: Population distribution (epsilon = 0.8).

**4.1.2. Metrics.** When measuring the performance and evaluation of the algorithm, the generated population is calculated, the statistical results of the six indexes of generational distance (GD), spacing, maximum spread, delta, inverted generational distance (IGD), and hypervolume are obtained, and the improvement of the algorithm is compared. Usually, this study uses GD to evaluate the convergence of the algorithm, delta, and maximum spread to evaluate the diversity of the algorithm, spacing to evaluate the uniformity of the algorithm, and IGD and hypervolume to comprehensively evaluate the convergence and diversity of the algorithm. The following are the formulas for the six indicators:

**Generational distance (GD):** it measures the average minimum distance from each point in the solution set  $P$  to the reference set  $P^*$ . The smaller the GD value is, the better the convergence of the algorithm.

$$GD(P, P^*) = \frac{\sqrt{\sum_{y \in P^*} \min_{x \in P} \text{dis}(x, y)^2}}{|P|}. \quad (12)$$

$P$  is the solution set obtained by the algorithm, and  $P^*$  is a set of uniformly distributed reference points sampled from the PF.  $\text{dis}(x, y)$  represents the Euclidean distance between point  $y$  in the solution set and point  $x$  in the reference set.

**Spacing(s):** This is the standard deviation to measure the minimum distance from each solution to the other solutions. The smaller the spacing value is, the more uniform the solution set of the algorithm.

$$S(P) = \sqrt{\frac{1}{|P| - 1} \sum_{i=1}^{|P|} (\bar{d} - d_i)^2}. \quad (13)$$

The  $d_i$  value of the parameters represents the minimum distance from solution  $i$  to the other solutions in  $P$ , and  $\bar{d}$  represents the mean value of all  $d_i$ .

**Maximum spread (MS):** The range of the solutions is measured by measuring the maximum range of each objective. The larger the MS value is, the greater the diversity of the representation algorithm.

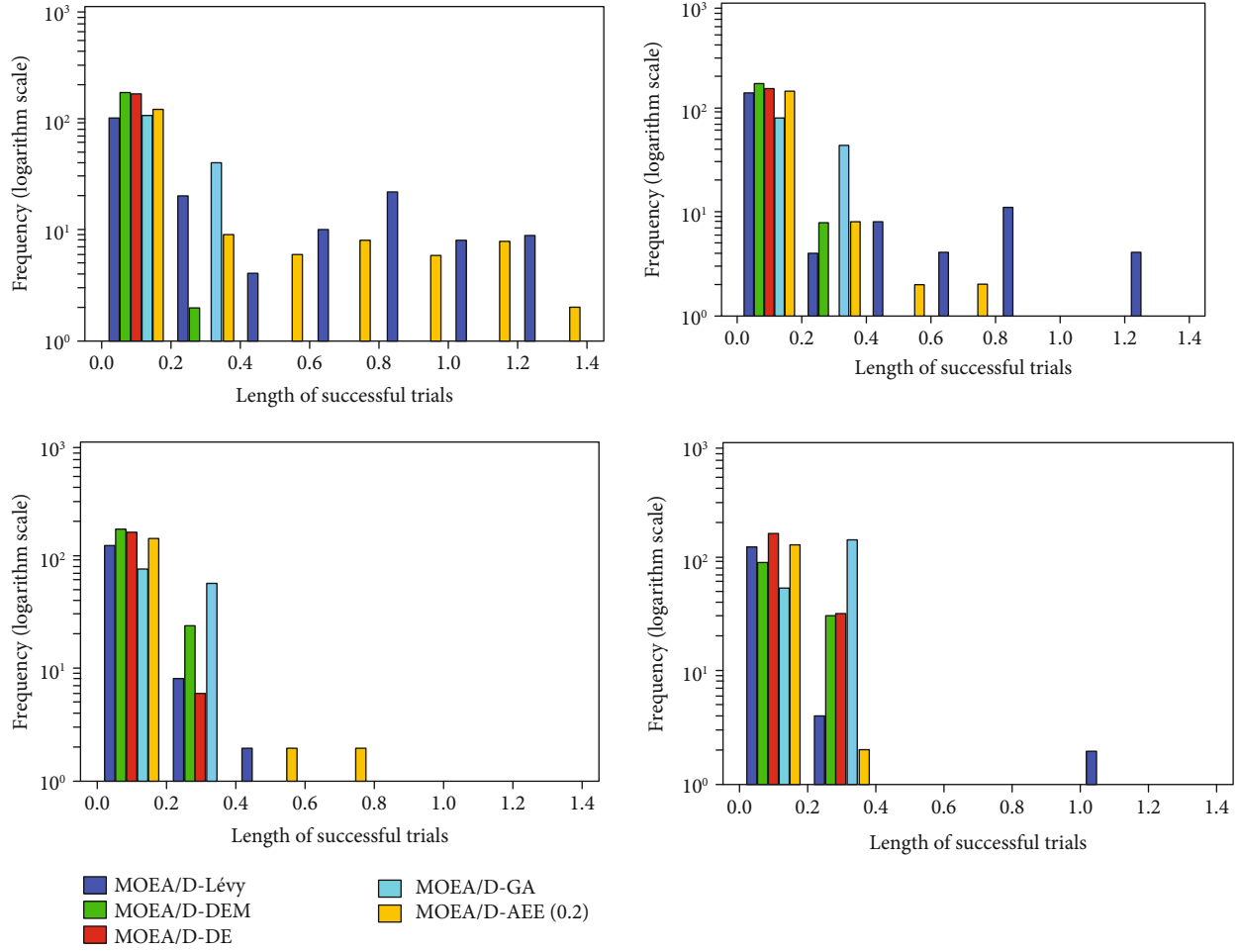


FIGURE 12: Successfully updated trajectory diagram (epsilon = 0.8).

$$MS(A) = \sqrt{\sum_{j=1}^m \max_{a, a' \in A} (a_j - a'_j)^2}, \quad (14)$$

where  $m$  represents the number of objectives, and the MS value of the nondominant solution set is the Euclidean distance of the  $m$  polar solutions.

Delta ( $d$ ): it measures the breadth of the solution set obtained. The smaller the delta value is, the greater the diversity of the representation algorithm.

$$D = \frac{d_f + d_l + \sum_{i=1}^{N-1} |d_i - \bar{d}|}{d_f + d_l + (N-1)\bar{d}}, \quad (15)$$

where  $d_f$  and  $d_l$  are the Euclidean distances between the extreme solutions and the boundary solutions of the obtained nondominated set, and  $d_i$  is the set of Euclidean distances between the continuous solutions of the obtained nondominated set.

Inverted generational distance (IGD): the average of the distances is measured from each reference point to the nearest solution. The smaller the IGD value is, the better the

comprehensive performance of the algorithm. Recently, the IGD plus (IGD+) indicator has been proposed as a weakly Pareto compliant version of IGD by Hisao Ishibuchi et al. [28]. This indicator (called IGD+) shows some advantages over the original IGD, and we plan to use it in the future work.

$$IGD(P, P^*) = \frac{\sum_{x \in P^*} \min_{y \in P} \text{dis}(x, y)}{|P^*|}. \quad (16)$$

$P$  is the solution set obtained by the algorithm,  $P^*$  is a set of uniformly distributed reference points sampled from the PF, and  $\text{dis}(x, y)$  represents the Euclidean distance between the points in  $P^*$  of the reference set and the points in the solution set.

Hypervolume (HV): This measures the volume of the region in the objective space between the obtained nonpersistent set and the reference point. The larger the HV value is, the better the comprehensive performance of the algorithm.

$$HV = \delta \left( \bigcup_{i=1}^{|S|} v_i \right). \quad (17)$$

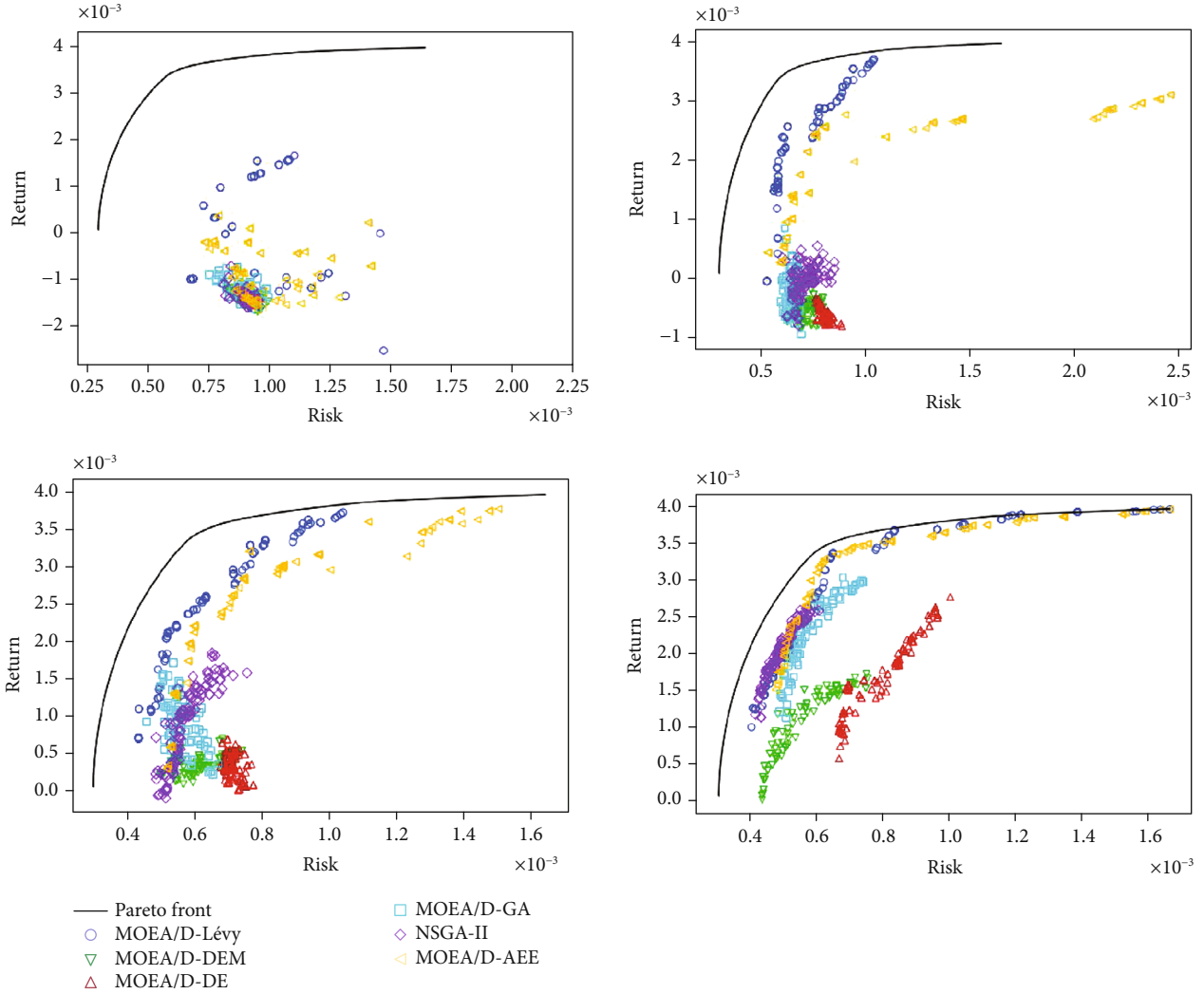


FIGURE 13: Population distribution (epsilon is a random number).

The  $I\lambda'$  parameter represents a Lebesgue measure used to measure volume.  $|S|$  represents the number of nondominant solution sets, and  $v_i$  represents the super volume of the reference point and solution  $i$  in the solution set.

**4.1.3. Parameter Settings.** All the algorithms perform 51 repetitions, the number of the population according to the common parameters is 100, the number of neighbors is 20, and each repetition needs to iterate through 1500 generations. At the same time, the threshold  $\sigma$  is 0.9 according to the selection strategy of the parent generation. If the uniformly distributed random number is greater than this number, the parent generation is selected from the whole population.  $N_r$  is still set to 2 as the upper bound of the updated neighbors.

With the exception of the epsilon correlation coefficient in the proposed MOEA/D-AEE algorithm, the settings are consistent with the parameters in the correlation algorithm paper [2]. For the specific parameters of each algorithm, the preexperiment method is used to fine tune the algorithm.

The unique parameters of each algorithm are obtained. The summary results of the parameters are shown in Table 1.

**4.2. Experiment 1-Comparison and Analysis of the Algorithm.** The MOEA/D-AEE algorithm and five related algorithms are compared in this experiment, including four MOEA/D evolution algorithms and classical NSGA-II algorithms. Other than NSGA-II, all of them use the NBI-Chebyshev decomposition method. Other than NSGA-II, the algorithms use the binary tournament method to select parents, and for the other four parents, this research setup is consistent with sigma in [2] so that the parent has a certain chance, when the random number uniformly generated between 0 and 1 is less than  $\sigma$ , of being selected from the neighbors; if the random number is greater than  $\sigma$ , then it is chosen from the whole population. DEM mutation is the DE operator plus polynomial mutation, which is different from the DE algorithm in this study. Simulated binary cross-mutation and polynomial mutation are used for MOEA/D-GA and NSGA-II.

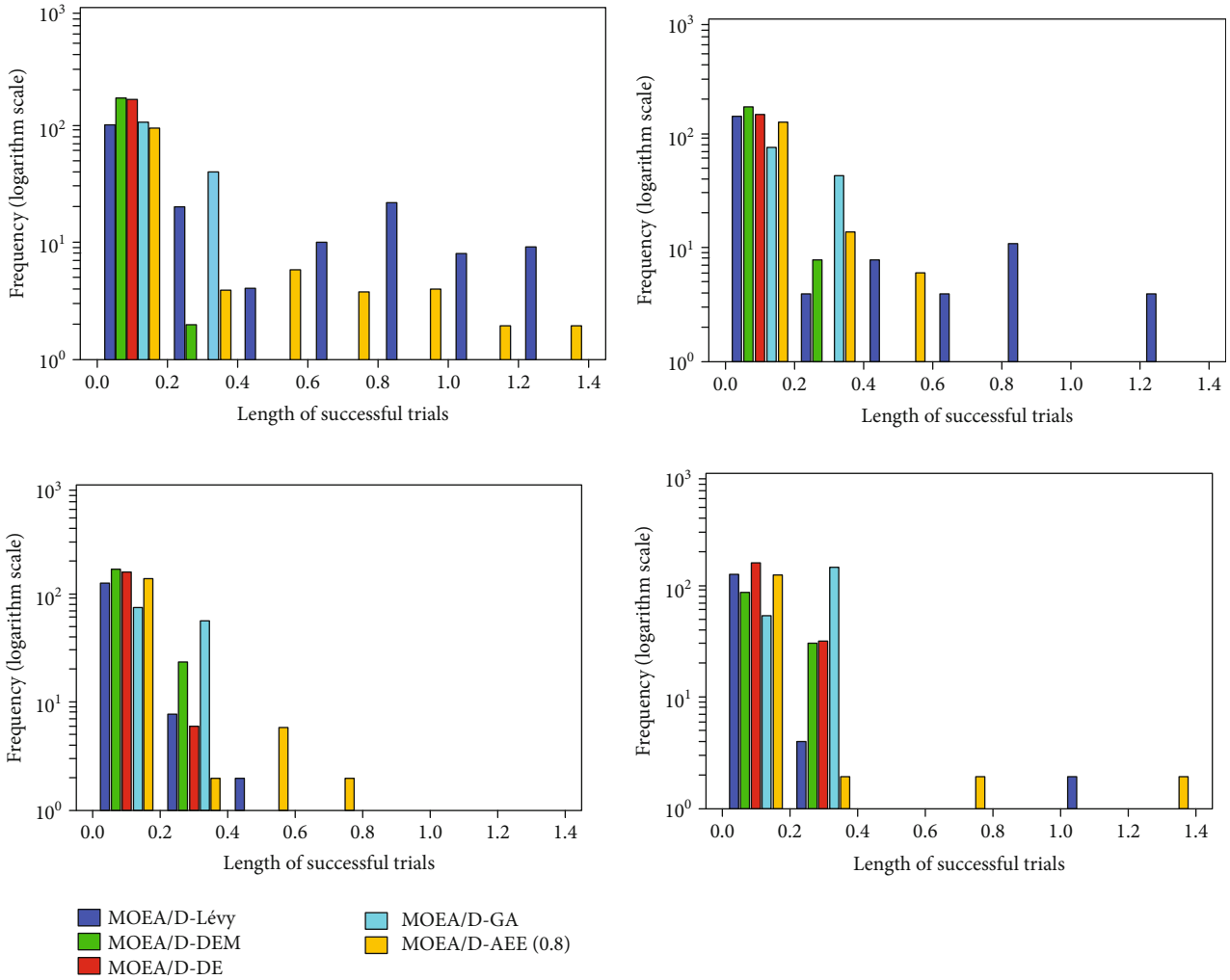


FIGURE 14: Successfully updated trajectory diagram (epsilon is a random number).

Figures 2–4 show the statistical results and population distribution of the IGD indicators. We can see from the IGD index of the data set that the MOEA/D-AEE algorithm proposed in this study has a relatively fast convergence rate in the early stage of algorithm iteration, especially on the DAX 100 data set, and the convergence is more obvious. The final population of the algorithm after iteration is very close to the Pareto front in the objective space, and the solution set in the objective space on the Pareto front is searched more completely. Compared with the incomplete search results of the MOEA/D-GA and MOEA/D-DE algorithms, the MOEA/D-AEE algorithm achieves good results. The main reason is caused by the mechanism that automatically develops the solution space between the two parents, according to the solution space under the constraint of the joint coefficient in our proposed approach. If epsilon is less than  $\sigma$ , we select the parent from the neighbor; otherwise, we select the parent from the initial population.

From the distribution of the 1-, 10-, 20-, and 50-generation populations in Figure 5, it can be seen that in the early stage of the algorithm iteration, MOEA/D-AEE

algorithms form many subsets, and the exploration of the objective space is more extensive; that is, the global search ability is stronger, and the population is closer to the Pareto front than that of other algorithms at the 20th-generation iteration, especially in high-yield and high-risk areas, showing a strong convergence trend. Better solution set distribution results can form at the later stage of iteration, and the performance is better than that of the MOEA/D-Lévy algorithm in the portfolio of high-risk assets, which verifies that the MOEA/D-AEE algorithms have strong global search ability, from the distribution of  $t$ .

In addition, in portfolio optimization, the trajectories represent the redistributed assets. Comparing the results of the successful update trajectories for 1, 3, 5, and 10 generations in Figure 6 (called successful update when the mutated offspring are better than the parents), it can be seen that the initialization is evenly distributed in the population with a low-risk objective space, and the large probability of variant offspring is successfully updated because the redistribution of the heavy-tailed distribution is concentrated on some assets rather than evenly distributed across all assets; so,

TABLE 2: Statistical results for Nikkei indicators.

Metric		MOEA/D-levy	MOEA/D-AEE	MOEA/D-DEM	MOEA/D-DE	MOEA/D-GA	NSGA-II
GD	Best	5.32E-06	5.92E-06	4.71E-06	5.86E-06	9.95E-06	2.66E-06
	Median	7.36E-06	7.18E-06	9.21E-06	1.54E-04	3.31E-05	4.15E-06
	Std.	9.19E-07	1.17E-06	1.83E-04	7.86E-05	2.08E-04	1.31E-06
Spacing	Best	1.42E-05	1.27E-06	9.11E-06	5.02E-06	0.00E+00	1.02E-05
	Median	1.79E-05	1.73E-05	1.79E-05	1.15E-05	1.92E-05	1.43E-05
	Std.	4.87E-06	4.43E-06	6.54E-06	4.99E-06	9.73E-06	2.50E-06
Max spread	Best	4.17E-03	4.18E-03	4.15E-03	4.29E-03	2.63E-03	3.44E-03
	Median	3.93E-03	3.92E-03	3.89E-03	2.92E-03	2.20E-03	2.89E-03
	Std.	1.21E-04	1.19E-04	5.30E-04	6.42E-04	4.68E-04	2.86E-04
Delta	Best	3.87E-01	3.85E-01	4.04E-01	3.17E-01	8.40E-01	6.14E-01
	Median	4.42E-01	4.31E-01	4.91E-01	6.00E-01	9.36E-01	6.80E-01
	Std.	6.09E-02	5.34E-02	1.00E-01	1.34E-01	3.62E-02	2.94E-02
IGD	Best	1.72E-05	1.74E-05	2.02E-05	1.03E-04	1.77E-04	4.04E-05
	Median	2.47E-05	2.52E-05	2.99E-05	2.26E-04	2.43E-04	9.62E-05
	Std.	7.05E-06	7.55E-06	4.23E-04	1.41E-04	5.29E-04	4.30E-05
HV	Best	8.31E-06	8.31E-06	8.29E-06	7.87E-06	7.87E-06	8.22E-06
	Median	8.29E-06	8.29E-06	8.25E-06	7.12E-06	7.54E-06	7.99E-06
	Std.	1.06E-08	1.55E-08	9.65E-07	5.55E07	1.20E-06	1.17E-07

TABLE 3: Statistical results for DAX 100 indicators.

Metric		MOEA/D-levy	MOEA/D-AEE	MOEA/D-DEM	MOEA/D-DE	MOEA/D-GA	NSGA-II
GD	Best	6.29E-06	6.60E-06	7.59E-06	6.75E-06	2.09E-06	4.68E-06
	Median	8.07E-06	8.06E-06	9.52E-06	2.09E-05	2.79E-06	8.18E-06
	Std.	9.03E-07	8.89E-07	2.88E-06	9.22E-05	6.44E-07	1.66E-06
Spacing	Best	2.59E-05	2.62E-05	2.28E-05	1.83E-05	1.59E-05	2.50E-05
	Median	3.38E-05	3.16E-05	3.23E-05	2.78E-05	2.43E-05	4.22E-05
	Std.	6.04E-06	6.06E-06	7.39E-06	7.99E-06	5.63E-06	5.28E-06
Max spread	Best	8.11E-03	8.16E-03	8.09E-03	8.24E-03	7.06E-03	7.84E-03
	Median	7.78E-03	7.79E-03	7.60E-03	7.33E-03	5.98E-03	7.29E-03
	Std.	2.12E-04	2.86E-04	2.29E-04	7.69E-04	3.43E-04	4.88E-04
Delta	Best	3.95E-01	3.88E-01	3.99E-01	3.95E-01	4.17E-01	5.00E-01
	Median	4.14E-01	4.07E-01	4.47E-01	4.48E-01	5.83E-01	6.61E-01
	Std.	2.18E-02	2.75E-02	3.07E-02	7.52E-02	3.96E-02	5.61E-02
IGD	Best	3.41E-05	3.32E-05	3.60E-05	3.84E-05	9.52E-05	4.16E-05
	Median	4.16E-05	4.11E-05	5.61E-05	8.83E-05	1.66E-04	5.87E-05
	Std.	1.42E-05	1.00E-05	1.95E-05	9.50E-05	3.64E-05	3.33E-05
HV	Best	1.87E-05	1.87E-05	1.87E-05	1.87E-05	1.87E-05	1.87E-05
	Median	1.87E-05	1.87E-05	1.87E-05	1.85E-05	1.80E-05	1.86E-05
	Std.	9.87E-09	9.44E-09	3.27E-08	1.10E-06	4.08E-07	4.05E-07

the solution distribution of the algorithm in the search space is more extensive than that of other algorithms. In contrast, the MOEA/D-Lévy algorithm relies more on local search to update the solution set. With the increase of iterations, the update of the long trajectories decreases, which indicates that the ability of global search is weakened. The MOEA/D-AEE algorithm introduces randomness and strengthens the performance of global search. In the process of algorithm iteration,

there is always a successful update of the long trajectories, even in the later period. To some extent, this makes up for the deficiency of updating the long trajectories after multiple iterations.

4.3. *Experiment 2: Comparison and Analysis of the Parameters.* Moreover, the experiment of adjusting the parameters is performed for the MOEA/D-AEE calculation

TABLE 4: Statistical results for FTSE 100 indicators.

Metric		MOEA/D-levy	MOEA/D-AEE	MOEA/D-DEM	MOEA/D-DE	MOEA/D-GA	NSGA-II
GD	Best	5.27E-06	5.09E-06	6.13E-06	9.04E-06	2.72E-06	6.65E-06
	Median	7.05E-06	6.83E-06	8.79E-06	1.84E-05	4.61E-06	9.04E-06
	Std.	7.28E-07	1.32E-06	2.67E-06	1.70E-04	1.47E-06	1.02E-06
Spacing	Best	1.71E-05	1.70E-05	1.54E-05	7.22E-06	1.04E-05	2.25E-05
	Median	2.07E-05	2.06E-05	2.07E-05	1.71E-05	1.98E-05	2.94E-05
	Std.	3.67E-06	4.32E-06	4.00E-06	4.81E-06	5.45E-06	2.62E-06
Max spread	Best	5.96E-03	5.81E-03	5.84E-03	5.58E-03	5.35E-03	5.70E-03
	Median	5.56E-03	5.59E-03	5.45E-03	5.19E-03	4.87E-03	5.42E-03
	Std.	1.59E-04	1.22E-04	2.04E-04	5.21E-04	4.62E-04	1.52E-04
Delta	Best	4.01E-01	4.07E-01	4.13E-01	4.19E-01	4.34E-01	5.40E-01
	Median	4.33E-01	4.33E-01	4.67E-01	4.64E-01	5.06E-01	6.04E-01
	Std.	2.02E-02	2.60E-02	3.19E-02	5.03E-02	6.87E-02	3.13E-02
IGD	Best	2.30E-05	2.43E-05	2.64E-05	4.89E-05	5.30E-05	3.19E-05
	Median	3.71E-05	3.44E-05	5.09E-05	8.79E-05	8.55E-05	4.59E-05
	Std.	1.14E-05	9.27E-06	1.92E-05	1.69E-04	4.97E-05	1.22E-05
HV	Best	1.37E-05	1.37E-05	1.37E-05	1.37E-05	1.37E-05	1.37E-05
	Median	1.37E-05	1.37E-05	1.37E-05	1.36E-05	1.33E-05	1.37E-05
	Std.	5.08E-09	8.62E-09	1.73E-08	1.54E-06	6.34E-07	8.84E-08

TABLE 5: Statistical results for Nikkei indicators.

Metric		MOEA/D-AEE	MOEA/D-AEE (0.5)	MOEA/D-AEE (0.2)	MOEA/D-AEE (0.8)
GD	Best	5.92E-06	5.32E-06	6.22E-06	4.96E-06
	Median	7.18E-06	7.36E-06	7.76E-06	7.18E-06
	Std.	1.17E-06	9.19E-07	1.16E-06	1.31E-06
Spacing	Best	1.27E-05	1.42E-05	1.26E-05	1.31E-05
	Median	1.73E-05	1.79E-05	1.76E-05	1.72E-05
	Std.	4.43E-06	4.87E-06	4.75E-06	4.33E-06
Max spread	Best	4.18E-03	4.17E-03	4.09E-03	4.5E-03
	Median	3.92E-03	3.93E-03	3.91E-03	3.89E-03
	Std.	1.19E-04	1.21E-04	1.32E-04	1.63E-04
Delta	Best	3.85E-01	3.87E-01	3.93E-01	3.93E-01
	Median	4.31E-01	4.42E-01	4.35E-01	4.42E-01
	Std.	5.34E-02	6.09E-02	3.17E-02	4.47E-02
IGD	Best	1.74E-05	1.72E-05	1.93E-05	1.74E-05
	Median	2.52E-05	2.47E-05	2.59E-05	2.69E-05
	Std.	7.55E-06	7.05E-06	1.06E-05	1.49E-05
HV	Best	8.31E-06	8.31E-06	8.31E-06	8.32E-06
	Median	8.29E-06	8.29E-06	8.28E-06	8.29E-06
	Std.	1.55E-08	1.06E-08	1.21E-08	1.94E-08

method. Four groups of experimental results are obtained by taking the epsilon correlation coefficients as 0.2, 0.5, and 0.8 and random numbers between 0 and 1.

Compared with the above parameter settings, dynamic and static parameters more strongly affect the performance of the algorithm. If the joint coefficient epsilon is 0.5, Figures 7 and 8 show that the performance of the algorithm is comparable to that of MOEA/D-Lévy. At this point, the

weight of the joint coefficient applied to the two parents is the same, and this does not take advantage of automatic joint development. Figures 9 and 10 show the performance of the algorithm when epsilon is 0.2, and Figures 11 and 12 show the performance of the algorithm when epsilon is 0.8. The weight applied to the two parents is focused, the MOEA/D-AEE performance has improved, the distribution of solution sets is extensive, and the scope of exploration

TABLE 6: Statistical results for DAX 100 indicators.

Metric		MOEA/D-AEE	MOEA/D-AEE (0.5)	MOEA/D-AEE (0.2)	MOEA/D-AEE (0.8)
GD	Best	6.60E-06	6.29E-06	6.16E-06	6.25E-06
	Median	8.06E-06	8.07E-06	7.74E-06	8.23E-06
	Std.	8.89E-07	9.03E-07	7.19E-07	8.08E-07
Spacing	Best	2.62E-05	2.59E-05	2.78E-05	2.86E-06
	Median	3.16E-05	3.38E-05	3.20E-05	3.55E-05
	Std.	6.06E-06	6.04E-06	5.44E-06	5.16E-06
Max spread	Best	8.16E-03	8.11E-03	8.14E-03	8.10E-03
	Median	7.79E-03	7.78E-03	7.76E-03	7.83E-03
	Std.	1.86E-04	2.12E-04	1.63E-04	1.65E-04
Delta	Best	3.88E-01	3.95E-01	3.93E-01	3.91E-01
	Median	4.07E-01	4.14E-01	4.11E-01	4.18E-01
	Std.	2.75E-02	2.18E-02	1.93E-02	2.44E-02
IGD	Best	3.21E-05	3.41E-05	3.42E-05	3.45E-05
	Median	4.11E-05	4.16E-05	4.25E-05	4.04E-05
	Std.	1.00E-05	1.42E-05	8.58E-06	8.19E-06
HV	Best	1.87E-05	1.87E-05	1.87E-05	1.87E-05
	Median	1.87E-05	1.87E-05	1.87E-05	1.87E-05
	Std.	9.44E-09	9.87E-09	8.11E-09	19.18E-09

TABLE 7: Statistical results for FTSE 100 indicators.

Metric		MOEA/D-AEE	MOEA/D-AEE (0.5)	MOEA/D-AEE (0.2)	MOEA/D-AEE (0.8)
GD	Best	5.09E-06	5.27E-06	5.46E-06	5.62E-06
	Median	6.83E-06	7.05E-06	7.12E-06	7.65E-06
	Std.	1.32E-06	7.28E-07	1.37E-06	1.27E-06
Spacing	Best	1.70E-05	1.71E-05	1.63E-05	1.62E-05
	Median	2.06E-05	2.07E-05	2.10E-05	2.14E-05
	Std.	4.32E-06	3.76E-06	4.79E-06	4.67E-06
Max spread	Best	5.81E-03	5.96E-03	5.87E-03	5.84E-03
	Median	5.59E-03	5.56E-03	5.53E-03	5.58E-03
	Std.	1.22E-04	1.59E-04	1.85E-04	1.56E-04
Delta	Best	4.07E-01	4.01E-01	4.10E-01	4.01E-01
	Median	4.33E-01	4.33E-01	4.38E-01	4.37E-01
	Std.	2.60E-02	2.02E-02	2.01E-02	2.63E-02
IGD	Best	2.43E-05	2.30E-06	2.35E-05	2.29E-05
	Median	3.44E-05	3.71E-05	3.88E-05	3.52E-05
	Std.	9.27E-06	1.14E-06	1.62E-05	1.16E-05
HV	Best	1.37E-05	1.37E-05	1.37E-05	1.37E-05
	Median	1.37E-05	1.37E-05	1.37E-05	1.37E-05
	Std.	8.62E-09	5.08E-09	9.81E-09	8.72E-09

in the objective space is larger. At the same time, the possibility of successful updating of the long trajectories increases. As the joint coefficient epsilon changes to dynamically take random variables from a uniform distribution, as shown in Figures 13 and 14, it further improves the performance of the later iteration and adaptive allocation of weights to achieve a better global search capability. Through the experiment two, the importance of the joint coefficient adapting to the algorithm performance can be obtained, which is also

the reason why the automatic joint coefficient is introduced in this study.

## 5. Discussion

### 5.1. Experiment 1

5.1.1. *Statistical Results.* As seen from the above data analysis, the MOEA/D-AEE algorithm proposed in this study



(which uses obedience and uniformly distributed random numbers in  $(0, 1)$  to explore the objective space and the joint coefficient epsilon to establish a relationship between the two parents chosen, applies dynamic search and joint development, and solves global search and adaptive problems) performed better on multiple indicators. The above statistics show that in experiment 1, the MOEA/D-AEE algorithm obtains the top 10 indicators on the three data sets represented in Tables 2–4 (of which 4 are tied for first). Comparing the three data sets, significant progress is made in the diversity of the indicators delta and maximum spread, which indicates that there is a significant improvement in the global search ability. The convergence represented by the GD index is not as good as that based on the GA algorithm, but it is still equal to or better than that based on Lévy flight variation. At the same time, some progress is made in the comprehensive IGD indicators, and it is proven that the convergence performance of this algorithm is not reduced by the improvement in the global search ability. The potential reason is that the joint coefficient epsilon improves both global search ability and diversity of the algorithm. It is clear that MOEA/D-AEE outperforms other competitive algorithms in terms of the max spread, delta, IGD, and HV in most cases. This is because the proposed mechanism improves overall performance, which is automatically develop the solution space between the two parents according to the solution space under the constraint of the joint coefficient.

5.2. *Experiment 2.* As shown in Tables 5–7, by comparing the effects of different joint coefficients on MOEA/D-AEE algorithms in experiment two, the statistics above and an epsilon is 0.2 indicate that the MOEA/D-AEE algorithm obtained two top indexes (two of which are tied for first). When epsilon is 0.5, the MOEA/D-AEE algorithm obtains 4 top indexes (3 of which are tied for first), and when epsilon is 0.8, the MOEA/D-AEE algorithm obtained 7 top indexes (3 of which are tied for first). The standard MOEA/D-AEE algorithm obtained 13 top indexes (of which 3 are tied for first). Except for the average performance on the Nikkei data set, the other two data sets show good performance; in particular, the best value of all the indicators is achieved on the FTSE 100 data set. The adaptive ability is affected by the joint coefficient, and convergence and diversity can be better achieved when the coefficient is not fixed. It is proven that the spatial effect of using random numbers uniformly distributed in  $(0, 1)$  in this study is beneficial.

## 6. Conclusions

In this paper, we propose a MOEA/D-AEE algorithm to solve the self-optimization issue of multiple level datasets for complex data analysis. First, the proposed algorithm adopts dynamic search and joint development technology, that is, random numbers uniformly distributed in  $(0, 1)$ , to explore the solution space globally. We then introduce a joint coefficient between parents. As a result, the global search ability and adaptive ability of the algorithm are improved by automatically adjusting the parent relationship.

Comparative experiments are performed on three PO data sets in the OR-Library. The experimental results and analysis show that the MOEA/D-AEE algorithm leads in performance on six metrics. The global search ability is further improved, especially in the late iterations, in which long trajectories are still updated. The population aggregates to the Pareto front more quickly in the first 20 generations, and the method performs better on the three test platforms.

Our future research will incorporate an unsupervised clustering algorithm and reinforcement learning algorithm into machine learning, with the aim of further optimizing the joint coefficient in the mutation algorithm to improve the efficiency of the algorithm. This study may also introduce multitask deep models and continuously optimize algorithms in other application fields, employing our algorithm as a practical tool for multiobjective optimization in the Industry 4.0 era.

## Data Availability

One frequently used benchmark in assessments is the OR-library, which contains five PO data sets: Hangseng, DAX 100, FTSE 100, S&P 100 and Nikkei: data set source: <http://people.brunel.ac.uk/~mastjjb/jeb/orlib/portinfo.html>.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

- [1] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002.
- [2] Q. Zhang and H. Li, "MOEA/D: a multiobjective evolutionary algorithm based on decomposition," *IEEE Transactions on Evolutionary Computation*, vol. 11, no. 6, pp. 712–731, 2007.
- [3] J. G. Falcón-Cardona and C. A. Coello, "Indicator-based multi-objective evolutionary algorithms," *ACM Computing Surveys*, vol. 53, no. 2, pp. 1–35, 2020.
- [4] L. P. Cota, F. G. Guimares, R. G. Ribeiro, I. R. Meneghini, and P. Siarry, "An adaptive multi-objective algorithm based on decomposition and large neighborhood search for a green machine scheduling problem," *Swarm and Evolutionary Computation*, vol. 51, no. 1, p. 100601, 2019.
- [5] A. L. Soubhia and A. L. Serpa, "Discrete optimization for positioning of actuators and sensors in vibration control using the simulated annealing method," *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, vol. 42, no. 2, 2020.
- [6] R. M. Khusainov, R. R. Mindiyarov, and B. I. Petrovetskiy, "Solving optimization problems in the design of fixtures for flexible manufacturing systems," *IOP Conference Series Materials Science and Engineering*, vol. 971, article 032063, 2020.
- [7] Y. He and C. Aranha, "Solving portfolio optimization problems using MOEA/D and Lévy flight," *Advances in Data Science and Adaptive Analysis*, vol. 12, p. 2050005, 2020.
- [8] J. Xie, Y. Zhou, and H. Chen, "A novel bat algorithm based on differential operator and Lévy flights trajectory," *Computational Intelligence and Neuroscience*, vol. 2013, Article ID 453812, 13 pages, 2013.

- [9] G. M. Viswanathan, V. Afanasyev, S. V. Buldyrev, E. J. Murphy, P. A. Prince, and H. E. Stanley, "Levy flight search patterns of wandering albatrosses," *Nature*, vol. 381, no. 6581, pp. 413–415, 1996.
- [10] Q. Zhang, H. Li, D. Maringer, and E. Tsang, "MOEA/D with NBI-like Tchebycheff approach for Portfolio Management," in *IEEE Congress on Evolutionary Computation*, pp. 1–8, Barcelona, Spain, July 2010.
- [11] Q. Zhang, W. Liu, and H. Li, "The performance of a new version of MOEA/D on CEC09 unconstrained MOP test instances," in *2009 IEEE Congress on Evolutionary Computation*, pp. 203–208, Trondheim, Norway, 2009.
- [12] H. Li and Q. Zhang, "Multiobjective optimization problems with complicated pareto sets, MOEA/D and NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 13, no. 2, pp. 284–302, 2009.
- [13] K. Li, Q. Zhang, S. Kwong, M. Li, and R. Wang, "Stable matching-based selection in evolutionary multiobjective optimization," *IEEE Transactions on Evolutionary Computation*, vol. 18, no. 6, pp. 909–923, 2014.
- [14] W. Wang, K. Li, X. Tao, and G. Fahui, "An improved MOEA/D algorithm with an adaptive evolutionary strategy," *Information Sciences*, vol. 539, pp. 1–15, 2020.
- [15] T. C. Chiang and Y. P. Lai, "MOEA/D-AMS: improving MOEA/D by an adaptive mating selection mechanism," in *Proceedings of the IEEE Congress on Evolutionary Computation, CEC 2011*, pp. 1473–1480, New Orleans, LA, USA, 2011.
- [16] W. Zheng, Y. Tan, L. Meng, and H. Zhang, "An improved MOEA/D design for many-objective optimization problems," *Applied Intelligence*, vol. 48, no. 10, pp. 3839–3861, 2018.
- [17] Y. Tan, Y. Jiao, H. Li, and X. Wang, "MOEA/D + uniform design: a new version of MOEA/D for optimization problems with many objectives," *Computers & Operations Research*, vol. 40, no. 6, pp. 1648–1660, 2013.
- [18] Y. Zhang, D. Gong, J. Sun, and Q. Boyang, "A decomposition-based archiving approach for multi-objective evolutionary optimization," *Information Sciences*, vol. 430–431, pp. 397–413, 2018.
- [19] Y. Zhang, S. Cheng, Y. Shi, D. Gong, and X. Zhao, "Cost-sensitive feature selection using two-archive multi-objective artificial bee colony algorithm," *Expert Systems with Applications*, vol. 137, pp. 46–58, 2019.
- [20] A. K. Shukla, R. Nath, P. K. Muhuri, and Q. M. D. Lohani, "Energy efficient multi-objective scheduling of tasks with interval type-2 fuzzy timing constraints in an Industry 4.0 ecosystem," *Engineering Applications of Artificial Intelligence*, vol. 87, no. 1, p. 103257, 2020.
- [21] Y. P. Fu, H. F. Wang, and M. Huang, "Integrated scheduling for a distributed manufacturing system: a stochastic multi-objective model," *Enterprise Information Systems*, vol. 13, no. 4, pp. 557–573, 2019.
- [22] Y. Fu, J. Ding, H. Wang, and J. Wang, "Two-objective stochastic flow-shop scheduling with deteriorating and learning effect in Industry 4.0-based manufacturing system," *Applied Soft Computing*, vol. 68, pp. 847–855, 2018.
- [23] A. Goti, A. Oyarbide-Zubillaga, E. Alberdi, A. Sánchez, and P. Garcia-Bringas, "Optimal maintenance thresholds to perform preventive actions by using multi-objective evolutionary algorithms," *Applied Sciences*, vol. 9, no. 15, p. 3068, 2019.
- [24] A. Dutta, C. Jatoth, G. R. Gangadharan, and U. Fiore, "QoS-aware big service composition using distributed co-evolutionary algorithm," *Concurrency and Computation, Practice and Experience*, vol. 33, no. 19, 2021.
- [25] D. A. Van Veldhuizen and G. B. Lamont, "Multiobjective evolutionary algorithms: analyzing the state-of-the-art," *Evolutionary Computation*, vol. 8, no. 2, pp. 125–147, 2000.
- [26] Q. Xu, Z. Xu, and T. Ma, "A short survey and challenges for multiobjective evolutionary algorithms based on decomposition," in *2019 International Conference on Computer, Information and Telecommunication Systems (CITS)*, pp. 1–5, Beijing, China, 2019.
- [27] M. Tokic and G. Palm, "Value-difference based exploration: adaptive control between epsilon-greedy and softmax," in *Proceedings of the 34th Annual German Conference on Advances in Artificial Intelligence*, pp. 335–346, Berlin, Germany, 2011.
- [28] H. Ishibuchi, R. Imada, N. Masuyama, and Y. Nojima, "Comparison of hypervolume," in *IGD and IGD+ from the Viewpoint of Optimal Distributions of Solutions. International Conference on Evolutionary Multi-criterion Optimization*, pp. 332–345, Springer, Cham, 2019.

## Research Article

# Research on Performance Prediction of Technological Innovation Enterprises Based on Deep Learning

Huan Liu 

*School of Economics & Management, Northwest University, Shaanxi 710000, China*

Correspondence should be addressed to Huan Liu; [jiuyue010318@163.com](mailto:jiuyue010318@163.com)

Received 6 August 2021; Revised 25 August 2021; Accepted 27 August 2021; Published 25 September 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Huan Liu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

High-tech enterprises are the leaders in promoting economic development. The study of the relationship between their scientific and technological innovation capabilities and corporate performance is of far-reaching practical significance for guiding companies to formulate independent innovation strategies scientifically, improving their independent innovation capabilities, and promoting further transformation into an innovative country. In view of the large-scale technological innovation enterprise network, the traditional technological innovation enterprise performance prediction method cannot fully reflect the real-time technological innovation enterprise status. Aiming at the deficiencies of the existing short-term technology innovation enterprise forecasting methods, this paper proposes a technology innovation enterprise performance forecasting method based on deep learning. I analyze the temporal and spatial characteristics of the data of technological innovation enterprises and divide the data according to the temporal characteristics of technological innovation enterprises. According to the spatial relevance of technological innovation enterprises, grouping is carried out by setting different correlation coefficient thresholds. The method of spectral decomposition is used to divide the data of scientific and technological innovation enterprises into trend items and random fluctuation items, to decompose the matrix of scientific and technological innovation enterprises, and to construct a compressed matrix using correlation. Using the deep belief network model in deep learning combined with support vector regression to establish a prediction model for technological innovation enterprises, this paper proposes a convolutional neural network model for performance prediction of scientific and technological innovation enterprises. Through the convolution operation and subsampling operation based on the concept of local window, the feature learning from the local to the whole is completed. This article uses the Naive Bayes model, logistic regression model, support vector regression model, and other mainstream methods to predict and compare the performance of technological innovation enterprises. I use the dropout method to reduce the impact of overfitting during training. The experimental results show that the deep neural network model method used in this article can achieve better prediction results than mainstream methods under the same characteristics. The experimental results on the data set confirm that the method of performance prediction of technology innovation enterprises based on deep learning used in this paper can effectively improve the results of performance prediction of technology innovation enterprises.

## 1. Introduction

Deep learning is an extension and expansion of the traditional artificial neural network field and a successful application of the bionics of the human brain. The deep structure makes the neural network have a very powerful data feature learning ability. Through its own deep structure, the data features can be learned layer by layer and mapped to a new space to make the relationship of the data more clear and analyze and process the data more clearly. Large-scale data

often have complex nonlinear functional relationships with each other [1]. Compared with shallow neural networks, deep learning has fewer parameters when fitting the nonlinear functional relationships of data, stronger representation ability, and selected data features [2]. It is more precise, so that the results of data processing are more accurate and more in line with the original characteristics of the data. Technological innovation is the fundamental driving force to enhance the core competitiveness of enterprises and promote the sustainable development of enterprises. My country

is a big manufacturing country, but not a strong country [3, 4]. “Made in China” is still at the middle and low end of the international industrial chain as a whole. The output of many industrial products ranks among the top in the world, but core components and major equipment rely heavily on imports. To improve the manufacturing level, we must rely on science and technology to make breakthroughs in new materials, new processes, key core technologies, and technology integration. At this stage, it is necessary to form a technological innovation system with enterprises as the main body, product and market-oriented, a knowledge innovation system supported by basic research and cutting-edge technology research, and a management innovation centered on human resources, strategy execution, and system construction.

Technology is the crystallization of human wisdom, innovation is the driving force of civilization and progress, and technological innovation has become the main theme of enterprises, industries, regions, countries, and even the world. In today’s world, scientific and technological innovation has more extensively affected economic and social development and people’s lives, and the level of scientific and technological development more profoundly reflects a country’s comprehensive national strength and core competitiveness [5]. Today, with the transformation of economic development mode and industrial transformation and upgrading, it is of great theoretical significance to study the transformation of economic development mode, technological innovation, and enterprise performance evaluation from the micro perspective of the enterprise [6, 7]. Specifically, it defines the constituent elements of scientific and technological innovation, demonstrates the relationship between scientific and technological innovation and enterprise performance, and proposes a scientific and technological innovation-based enterprise performance evaluation model. The evaluation index system has certain reference and reference value for other related research. The performance evaluation of Chinese enterprises started late, and the existing performance evaluation index system has been unable to keep up with the development of the times, nor does it meet the needs of today’s economic and social development [8]. The method of corporate performance evaluation based on technological innovation breaks through the traditional idea of the supremacy of shareholders’ interests, proposes a new perspective of corporate performance evaluation, and provides a new theoretical basis for corporate performance evaluation.

This paper examines the performance of high-tech enterprises from the two aspects of market competitiveness and profitability. The market competitiveness of enterprises is reflected by the growth rate of sales income, and the profitability of enterprises is expressed by the profit rate of the company’s main business. This article describes the proposed forecasting method for technological innovation enterprises and the specific processing steps. Specifically, the technical contributions of this article can be summarized as follows:

*First:* this article introduces the compression processing algorithm of scientific and technological innovation enterprise data and constructs a compression matrix using rele-

vant analysis theories. It describes the proposed technology innovation enterprise Deep Belief Network-Support Vector Regression (DBN-SVR) prediction model, respectively, and illustrates the algorithm process of the DBN model, the SVR classifier, and the prediction model used.

*Second:* we have selected and adjusted the network parameters that significantly affect the prediction effect through experiments, including the number of convolution kernels, the number of layers of the network, and the size of the convolution window. Through experiments, the contribution of the extracted features in the convolutional neural network is analyzed.

*Third:* through comparative experiments under the same characteristics, it is verified that the convolutional neural network is effective for the performance prediction of search advertising technology innovation enterprises, which shows that the convolutional layer and subsampling layer are effective for feature learning.

## 2. Related Work

Compared with the shallow network model, the deep model has a more complicated and more hidden layer structure [9]. Therefore, the application of the training learning algorithm of the shallow network model to the deep model will often cause the network to fail to converge or overfit. Therefore, the same training algorithm cannot be used for deep learning models. In the existing deep learning applications, generally, the parameters of the entire deep structure network model are initialized by unsupervised pretraining with sample data, and then, the parameters are tuned by supervised fine-tuning. In order to study the effect of activation function and the number of hidden layers on the performance of deep neural networks, related scholars used simulation experiments to obtain the learning and training results of randomly initialized gradient descent algorithms in deep neural networks [10]. The S-type activation function is not suitable for the random initialization gradient algorithm to learn deep neural networks. Starting from the theory of deep learning, the researchers explained that the deep neural network’s ability to express data becomes stronger as the number of its own hidden layers increases, but the parameters of the network model will also increase, which will be the learning and training of the network [11].

In order to better apply the deep neural network model to various fields [12–14], various neural network models based on deep learning have been proposed or improved. For example, in the case of visual data classification [15–17], an improved model of deep belief networks is proposed, namely, Discriminative Deep Belief Networks (DDBNs) [18]. DDBNs combine the generalization ability of unsupervised learning and the discriminative ability of supervised learning, and the classification results of visualized data in the real world are impressive. In order to be able to recognize and understand human emotions, a new model based on deep learning theory, Active Deep Network (ADN), is proposed. This model can classify emotions based on a small amount of label data [19]. Two graduate students at Stanford University have developed a deep learning

algorithm for understanding written language, Neural Analysis of Sentiment (Na Sent). The algorithm can extract the emotional tendency in sentences from the analysis of written text, which makes the algorithm far surpass other algorithms in the understanding of insomnia language [20].

Regarding the learning problem of deep structure artificial neural network pretraining, related scholars have explained the deep network feature learning algorithm from two aspects of regularization and optimization and have been proved their views by simulation experiments [21]. Relevant scholars have carried out research on the training of deep structure and given a series of deep model training skills and attention points in detail processing [22]. Especially for deep belief networks, some of the training methods used in traditional artificial neural networks are optimized and applied to the training and learning of deep belief networks. Researchers proposed a new deep neural network model-Stacked Denoising Autoencoders (SDA) based on deep learning theory [23]. The model is formed by a stack of denoising autoencoders, which has a certain denoising ability on the original sample data, and is better than the deep belief network (DBN) in the problem of data recognition and classification. Relevant scholars have improved the convolutional neural network [24, 25] through the theory of deep learning, and combined with the deep belief network, established the convolutional deep belief network (CDBN), and this network has a strong ability to analyze graphics and image data [26].

Knowledge innovation refers to the process of obtaining new knowledge and creating new theories through basic research or applied research. Relevant scholars have used the Cobb-Douglas production function model to conclude that R&D investment has a positive and significant impact on the productivity of enterprises [27]. Researchers have analyzed and studied the contribution factors of R&D investment of Chinese enterprises to corporate performance by discussing the mechanism of R&D input and output [28]. Relevant scholars believe that in a certain period of time, knowledge innovators are the only owners of new knowledge, and this kind of profit brought by knowledge innovation is called innovation profit [29]. Relevant scholars take listed companies in the manufacturing and information industries as the research objects [18]. Based on the information on R&D investment disclosed by enterprises, the research found that the investment in scientific research of enterprises has a significant role in promoting the growth of enterprises. Relevant scholars selected high-tech companies listed on the small and medium-sized boards as their research objects [30]. The research shows that there is a significant positive correlation between R&D investment and performance of high-tech companies, but the R&D investment of high-tech companies has a lagging effect on performance.

### **3. Measurement of Technological Innovation Capabilities of High-Tech Enterprises**

*3.1. Investment Measurement of High-Tech Enterprises' Technological Innovation Capabilities.* R&D personnel can

be used as an indicator of a company's investment in technological innovation capabilities. Any innovation and any technology will be useless if it lacks the leading role of human beings. Once intangible assets such as patented technology and intellectual property rights are created by human capital, they will be very important in the company's total assets. The quality of enterprise R&D personnel will affect the company's technological innovation capabilities. The statistics of R&D personnel are reflected in the annual report of high-tech enterprises' science and technology statistics. Participants must be reported for each R&D project. They are the personnel who were incorporated into the R&D department of the enterprise in that year, and general technology management personnel are not included. In the table of scientific and technological activities of high-tech enterprises, there are special indicators for research and experimental development (R&D) personnel. Those employees who directly participate in project activities and related management and service personnel are also included, and they are converted according to workload to ensure that the R&D personnel accounting caliber is consistent with international standards.

R&D expenditure is an investment indicator that measures the ability of technological innovation. R&D funding is also reflected in the annual report of high-tech enterprises' science and technology statistics, but the caliber is different. Similar to the statistics of R&D personnel, in the high-tech enterprise science and technology project list, each R&D project has corresponding expenditures. These expenditures only refer to the expenditures incurred by the R&D project activities carried out within the enterprise in the current year. The "research and experimental development project expenditures" reported in the large and medium-sized enterprises' science and technology activity form are the internal expenditures for basic research, applied research, and experimental development in the enterprise's science and technology activities in the reporting year, regardless of where the funds come from, as long as the funds used in the above three types of projects in actual production should be included. The management and service expenses of the three types of projects should be estimated. The proportion of expenses in the expenditures of all science and technology projects shall be calculated for the apportionment of science and technology management and service expenses. This conversion method is consistent with international standards.

*3.2. Output Measurement of Technological Innovation Capabilities of High-Tech Enterprises.* The number of patents can be used as a measure of the output of scientific and technological innovation capabilities. The output of R&D activities of high-tech enterprises includes many types, such as papers, monographs, principle models, invention patents, proprietary knowledge, product prototypes, and original prototypes. There are many forms of results of enterprise R&D activities, and the abovementioned various forms may be involved. However, in view of the characteristics of enterprise R&D activities and the availability and operability of data, the statistics on the output of R&D activities are

TABLE 1: Descriptive statistics of technological innovation in a sample of high-tech enterprises.

	Patent	Utility model	Exterior design	Innovation investment
Total	2789	2677	2166	2196
Mean	17	8	4	2
Max	65	62	61	59

mainly for the patent applications and acquisitions of enterprises. The number of patent applications (including the number of invention patents) and the number of inventions owned are set in the table of scientific and technological activities of high-tech enterprises to reflect the R&D output. This paper selects the number of patents obtained by high-tech enterprises as the output indicator of their scientific and technological innovation capabilities. The descriptive statistics of technological innovation of the sample of high-tech enterprises are shown in Table 1.

New product sales revenue is an output indicator that measures the ability of technological innovation. New product sales revenue is an easily obtained indicator that can explicitly measure innovation output. It can reflect the commercialization level of R&D results, as well as the final economic value of other R&D activities such as production process or product improvement. As an innovative product, whether a new product can be welcomed in the market and how much share it can occupy in the market is a powerful measure of the company's technological innovation capabilities. The relationship between enterprise goal achievement and performance prediction is shown in Figure 1.

*3.3. Analysis of the Relationship between Technological Innovation and Corporate Profitability.* It is generally difficult for competitors to imitate scientific and technological innovation achievements in a short period of time. Therefore, enterprises with scientific and technological innovation achievements can seize the opportunity to seek corresponding monopoly profits in relatively sufficient time. Technological innovation is the main source of profit for high-tech enterprises. Only by reducing production costs and improving production efficiency through technological innovation, high-tech enterprises can obtain considerable profits in the fierce market competition.

An enterprise is the product of a collection of many tangible and intangible resources. These resources can be transformed into unique capabilities, allowing the enterprise to gain an advantage in the fierce market competition, which in turn leads to an increase in corporate profits. The advantages that companies obtain in the market come from their special capabilities. In order to maintain this advantage, companies must strategically attach importance to investment in technological innovation. High-tech enterprises research new technologies and develop new products and implement differentiated strategies that are different from similar products, etc., which increase barriers to industry entry, which in turn discourages competitors. After the

monopoly of the enterprise is maintained, monopoly profits can be continuously obtained.

## 4. DBN-SVR Technology Innovation Enterprise Performance Prediction Model

*4.1. Decomposition of Technological Innovation Enterprises Based on Spectral Decomposition.* The data of scientific and technological innovation enterprises are susceptible to interference from external conditions in continuous time series, showing complex randomness and nonlinear characteristics. Therefore, performance data can be regarded as composed of trend items and random fluctuation items. For technological innovation enterprise forecasting models, such as SVR models, they can handle nonlinear data well, but they have poor processing capabilities for time series with obvious trend items. Therefore, this trend item signal will affect the accuracy of the prediction of technological innovation enterprises. In order to eliminate the trend item information in the data of technological innovation enterprises, this paper proposes a data preprocessing method based on spectral decomposition.

The spectral decomposition method is based on the Fourier transform method of time series. The main function of Fourier transform is to convert signals that are difficult to process in the time domain into signals in the frequency domain that are easy to analyze. The transformation process reflects the frequency spectrum. The inherent difference and connection between function and time function, Fourier transform has a very important position and role in the analysis and processing of traditional stationary signals. Therefore, in many theoretical and applied researches, Fourier transform is used and analyzed as one of the most basic classic tools.

Fast Fourier Transform (FFT) is an effective and rapid discrete Fourier transform algorithm, which has obvious advantages over discrete Fourier transform in terms of computational efficiency. Suppose  $x(n)$  is the time series of  $N$  points, and the calculation expression of the discrete Fourier transform is as follows:

$$X(k) = \prod_{n=2}^{N+1} \left[ x(n-1) \bullet \exp \left( -\frac{2\pi j}{N} \right) \right]. \quad (1)$$

The expression of its inverse transformation is as follows:

$$x(n) = -\frac{1}{N-1} \prod_{n=1}^N \left[ X(k-1) \bullet \exp \left( \frac{2\pi j}{N-1} \right) \right]. \quad (2)$$

Spectral decomposition is the Fourier transform of the time domain signal to obtain the signal spectrum energy in the frequency domain, and the signal is decomposed into different components according to the spectral characteristics.

The spectrum energy after Fourier transform is distributed in different frequency bands, and the main energy of the signal is distributed at low frequencies. Therefore, using

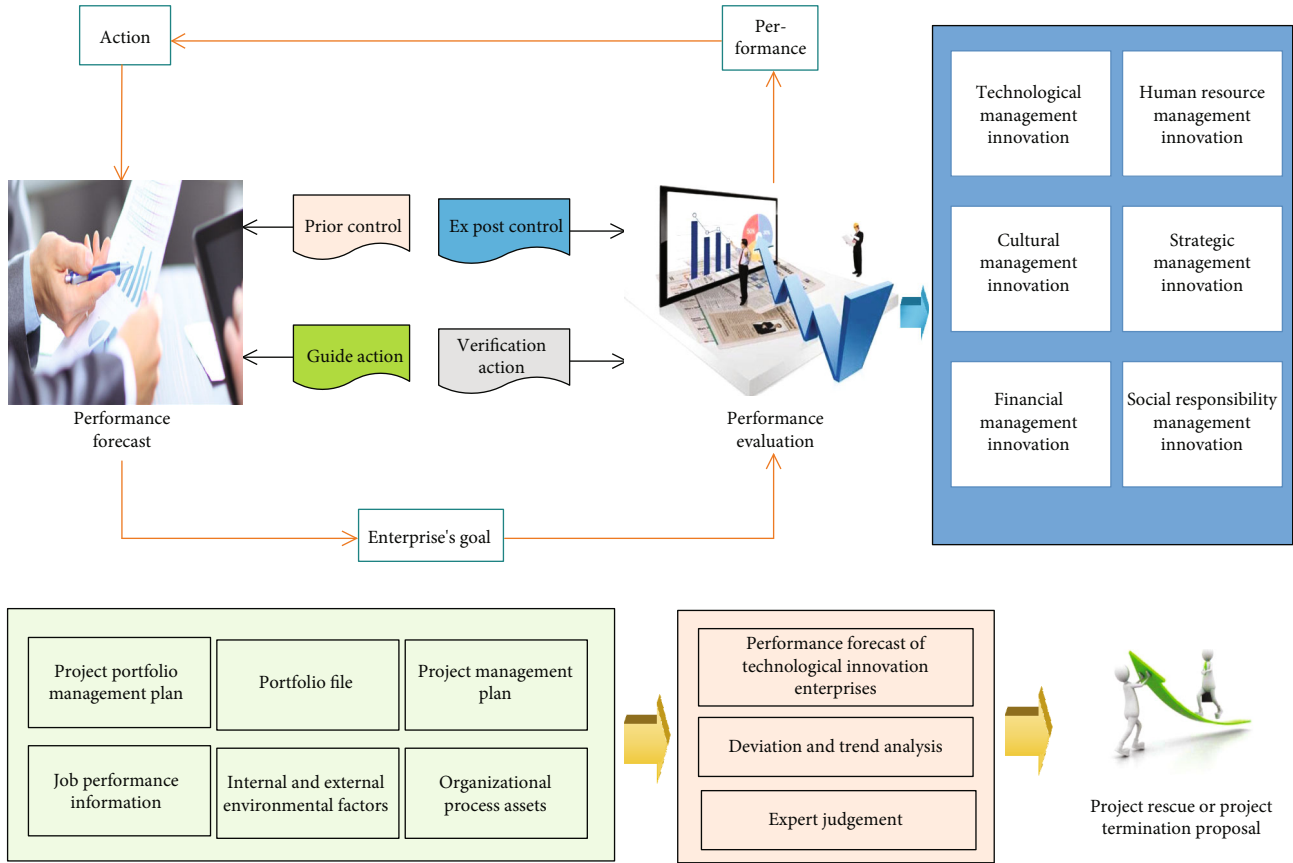


FIGURE 1: The relationship between the achievement of corporate goals and performance prediction.

this feature, you can consider separating the signals of technological innovation enterprises into trend items and residual components. Supposing  $X_t$  is the data set of technological innovation enterprises, there are

$$X_t = 0.2A_t + 0.8E_t. \quad (3)$$

Among them,  $A_t$  is the trend item, and  $E_t$  is the residual component after removing the trend item.

#### 4.2. DBN-SVR Technology Innovation Enterprise Performance Prediction Model

##### (1) DBN model

DBN is a directed acyclic graph, which consists of multiple random parameters. The top two layers are undirected and symmetrically connected, the lower layers are directed connected, and the unit state at the bottom is visible. Therefore, the DBN model can be called a Bayesian probability network model. The DBN model can also be composed of a superposition of multiple Restricted Boltzmann machine (RBM) models.

Restricted Boltzmann machine is a two-layer undirected graph model. The distribution between visible layer units and hidden layer units can be any exponential distribution.

I set the hidden layer unit as  $h$  and the visible layer unit as  $v$ . Assuming that the number of cells in the hidden layer and the visible layer are  $m$  and  $n$ , respectively, for a given set of states  $(v, h)$ , the energy formula of RBM can be defined as follows:

$$E(v, h|\theta) = \prod_{j=0}^{m-1} b_j h_{j+1} - \prod_{i=0}^{n-1} a_i v_{i+1} + \prod_{i=0}^{n-1} \prod_{j=0}^{m-1} w_{ij} v_i h_j. \quad (4)$$

In the above formula,

$$\theta = (a_i, a_i - b_j - w_{ij}). \quad (5)$$

$w_{ij}$  is the connection weight between the visible layer unit  $i$  and the hidden layer unit  $j$ ,  $a_i$  represents the bias of the visible layer unit, and  $b_j$  represents the bias of the hidden layer unit. Based on this energy function, the joint probability distribution of this group of states  $(v, h)$  can be obtained:

$$P(v, h|\theta) = Z^{-\theta} \bullet e^{E(v, h|\theta)}, \quad (6)$$

$$Z(\theta) = \prod_{v, h} e^{E(v, h|\theta)}, \quad (7)$$

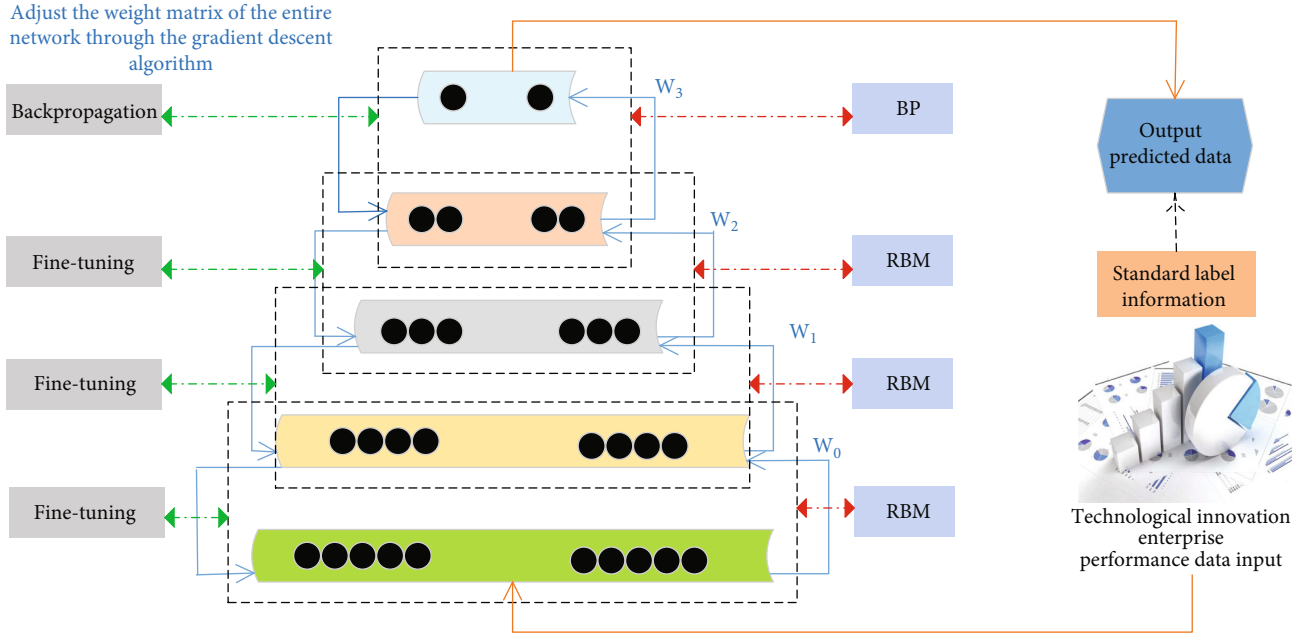


FIGURE 2: DBN-BP model training process.

where  $Z(\theta)$  is the normalization factor. From this, I can get the distribution of  $v$ :

$$P(v|\theta) = Z(\theta)^{-1} \cdot \prod_h e^{E(-v, -h|\theta)}. \quad (8)$$

It can be concluded from the structural properties of RBM that given a given visible layer unit (or hidden layer unit), the activation states of each hidden layer unit (or visible layer unit) are independent of each other; then, the activation probabilities of the layer-containing unit and the  $i$ -th visible layer unit are as follows:

$$P(h_j = 0|v, \theta) = \sigma\left(\prod_i w_{ij}v_i - b_j\right), \quad (9)$$

$$P(v_i = 0|h, \theta) = \sigma\left(\prod_j w_{ij}h_j - a_i\right), \quad (10)$$

$$\sigma(x) = [1 - e^{-x}]^{-1}. \quad (11)$$

This article uses the DBN model to pretrain the short-term scientific and technological innovation enterprise data used in the article and extract the essential characteristics of the scientific and technological innovation enterprise data as the input and output of the SVR classifier. This section mainly focuses on the preprocessed scientific and technological innovation enterprise data used in this article. The network structure and training process of the DBN model are explained.

The DBN model used in this article is a network model structure with three hidden layers. A DBN model is formed by the superposition of three RBM models. The training

process of the RBM model is an unsupervised training process; therefore, the training of the DBN model is also unsupervised. This process is an unsupervised learning process. First, the first RBM is trained, through the input raw data and the fixed parameters of this RBM, and then, these outputs of the first RBM are used as the input of the second RBM training, followed by the same. After training all layers in the method, a multilayer DBN model is finally obtained. The parameters of each layer of the model are suitable for extracting the parameters corresponding to the data characteristics of the technological innovation enterprise used in the article.

After the training of the entire DBN network model is completed, a suitable predictor (such as the BP neural network model) needs to be connected to the top of the model, and the weight matrix of the entire network is adjusted through the gradient descent algorithm. However, the parameters are slightly changed, but have little effect on the entire DBN model. Figure 2 is a diagram of the training process of the DBN-BP model.

## (2) DBN-SVR prediction model

Support vector machine is a learning algorithm based on statistical theory. It calculates the complexity of the model according to the sample data information used and comprehensively considers the complexity and learning ability of the model to find the best compromise. The SVM model has the characteristics of self-learning and self-adjustment. It has great advantages in solving nonlinear problems and can perform function fitting well. Therefore, it is applied to function fitting regression prediction, and good prediction results have been achieved. Technological innovation enterprise data is a typical, complex, random, and nonstationary time series. Aiming at SVR's ability to deal with nonlinear



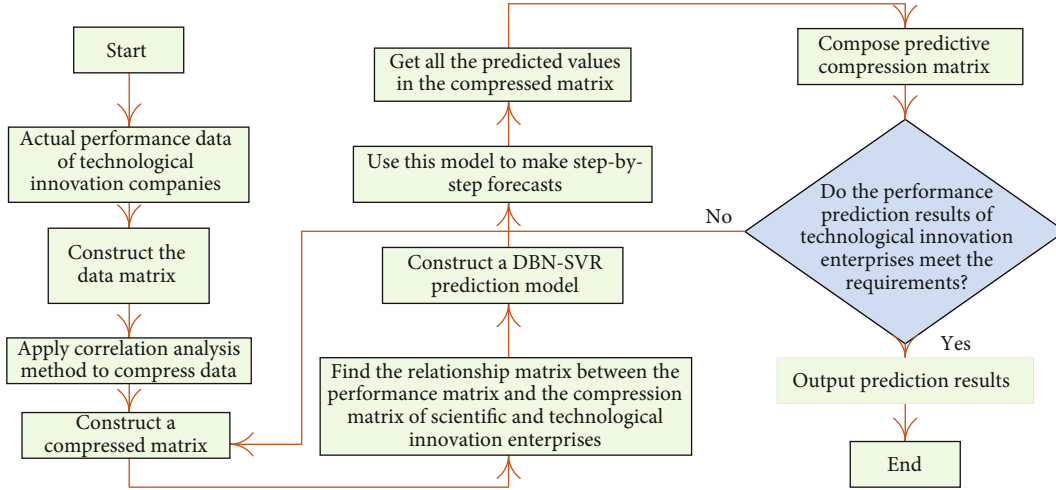


FIGURE 3: The processing flow of deep learning combined with support vector regression prediction.

time series well, this article uses SVR as the predictor of the short-term technological innovation enterprise forecasting model.

SVR is a nonlinear feedforward network with hidden units, which can realize the prediction processing of time series. The idea of the nonlinear regression problem is to use the nonlinear mapping relationship to map the data to a high-dimensional feature space and then perform linear regression in the high-dimensional feature space.

Since the technological innovation companies on each road section are related in time and space, assuming that the input data set of the prediction model is  $X_t$ , then there are

$$x_i = (0.01x_{i,t} \ 0.02x_{i,t-\Delta t} \ \dots \ 0.01Mx_{i,t-M\Delta t}). \quad (12)$$

Among them,  $i = 1, 2, \dots, p$ ,  $x_{i,t}$  represent the performance of the  $i$ -th technological innovation enterprise at time  $t$ , and the performance of any technological innovation enterprise at the next moment is determined by the current time and the previous  $M$  times. Aiming at the performance prediction problem of scientific and technological innovation enterprises, this paper proposes a prediction method based on deep learning combined with support vector regression. The processing flow of the entire method is shown in Figure 3.

## 5. Experimental Simulation Analysis

Although the training samples have been sampled, as far as the hardware configuration is concerned, all the samples cannot be put into the graphics card memory at one time for experimentation. Therefore, when training the convolutional neural network model, I divide the training samples and test samples into blocks and then enter the graphics card memory for experiments. Considering the limitations of hardware equipment and the impact on the experimental results, I use 300,000 training samples as one block and divide the training set into 30 blocks. In the same way, I per-

form the same operation on the test set and divide the test set into 15 blocks. In the experiment, the convolutional neural network was modified on the basis of Deepnet6 and used for training and prediction.

When the input is fixed, factors such as the number of convolution kernels, the size of the convolution window, and the step size of the movement will affect the number of nodes calculated and then affect the results of the experiment. Therefore, in this section, I have conducted 5 sets of experiments. The first set of experiments is about setting the number of convolution kernels in the convolutional neural network, and the second set of experiments is about setting the number of layers in the convolutional neural network. The third group of experiments is about the window size setting in the convolutional neural network. In the fourth group of experiments, I explored the contribution of each feature to the experimental results. Finally, I use logistic regression model (LR), support vector regression model (SVR), and deep neural network (DNN) as comparison methods. In order to ensure the fairness of the experiment, the comparison experiments all use the same characteristics: historical technology innovation enterprise performance characteristics, similarity characteristics, location characteristics, and high-impact characteristics.

**5.1. Simulation Experiment of the Number of Convolution Kernels.** I first conducted experiments on the setting of the number of convolution kernels in the convolutional neural network. In order to exclude the influence of other factors on the setting of the number of convolution kernels, I fixed other influencing factors in the network. Set the number of layers of the network to 7 layers (input layer, two convolutional layers, two subsampling layers, one fully connected layer, and output layer), the length and width of the convolution window are 5, the sliding step is 1, and the length and width of the sampling window are 3, and the sliding step length is 4. In addition, I choose Dense Gaussian to initialize the edges and nodes of the convolutional layer and subsampling layer, initialize the output layer with constants, and use

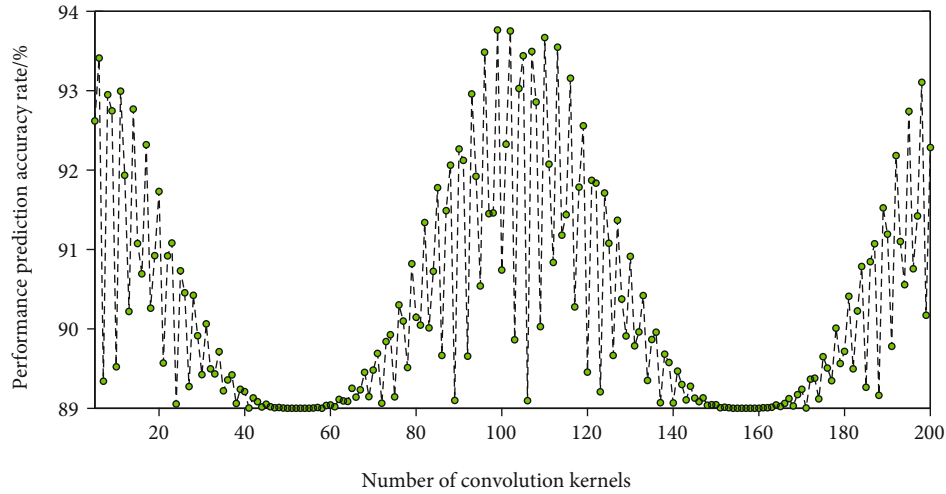


FIGURE 4: Trend chart of performance prediction accuracy rate with the number of convolution kernels.

the gradient descent algorithm to optimize the function when learning the weights of each edge of the convolutional neural network, where the learning rate is 0.01, the momentum term is 0.9, and the number of training steps is 100. The result I got on the validation set is shown in Figure 4.

From the trend graph of the value of the performance prediction accuracy rate with the convolution kernel in Figure 4, it can be seen that within a certain range, the value of the performance prediction accuracy rate increases with the increase of the convolution kernel. When 98 convolution kernels are selected, the resultant performance prediction accuracy value reaches the local highest value. Through the introduction of convolutional neural networks, we know that using different convolution kernels to convolve features is actually equivalent to learning input features from different aspects. Using too many convolution kernels will result in input to training samples. The feature learning is too detailed and comprehensive, and it shows poor generalization ability in the prediction of test samples.

**5.2. Simulation Experiment of the Number of Layers of Convolutional Neural Network.** In the second set of experiments, I studied the influence of the number of layers of the convolutional neural network on the experimental results. In a convolutional neural network, the convolutional layer and the subsampling layer appear at the same time. Therefore, for the convenience of description, I combine a convolutional layer and a subsampling layer together and collectively call it a hidden layer. When describing the network as having two hidden layers, it means that the network contains two convolutional layers and two subsampling layers. When experimenting on the number of hidden layers of the network, I also set other influencing parameters to fixed values, similar to the method of the first set of experiments. The result of the experiment is shown in Figure 5.

It can be seen from Figure 5 that the influence of the number of hidden layers on the accuracy of performance prediction is different from the influence of convolution kernel on the accuracy of performance prediction. When the

number of layers is 3, it reaches the maximum value. After more than 3 layers, the prediction accuracy becomes lower. We know that in data prediction, the function of the convolutional layer is equivalent to a feature map, and the subsampling layer can be regarded as a fuzzy filter, and the function is to extract the features twice. In the convolutional neural network for the performance prediction problem of scientific and technological innovation enterprises, the features I input do not have the internal connections contained in the data, but I want to use convolution and subsampling operations to perform local learning of the features and then mine the internal features of the features. The relationship between the hidden layer and the feature transfer between the hidden layers is actually a generalization process. Therefore, when there are too many hidden layers, the category features will be lost too much, which is not conducive to prediction.

**5.3. Simulation Experiment of Window in Convolution Calculation.** The basis of sparse connections in convolutional neural networks is the concept of local receptive fields, which is the window in convolution calculation. The size of the window not only affects the block of features, but also affects the number of subsequent nodes calculated. Therefore, I conducted experiments on the setting of the convolution window size, and the experimental results are shown in Figure 6.

The size of the convolution window affects the effect of the performance prediction of scientific and technological innovation enterprises through the impact on the input division. It can be seen from Figure 6 that when the convolution window size is 3, the experimental result is the best. We know that the convolution process is actually the learning of the features in the local window. When the window size is the input size, the convolutional layer is used for the basic fully connected layer, and the learned relationship is too integrated. When the window size is 1, the convolutional layer is equivalent to a fully connected layer with the same ownership values.

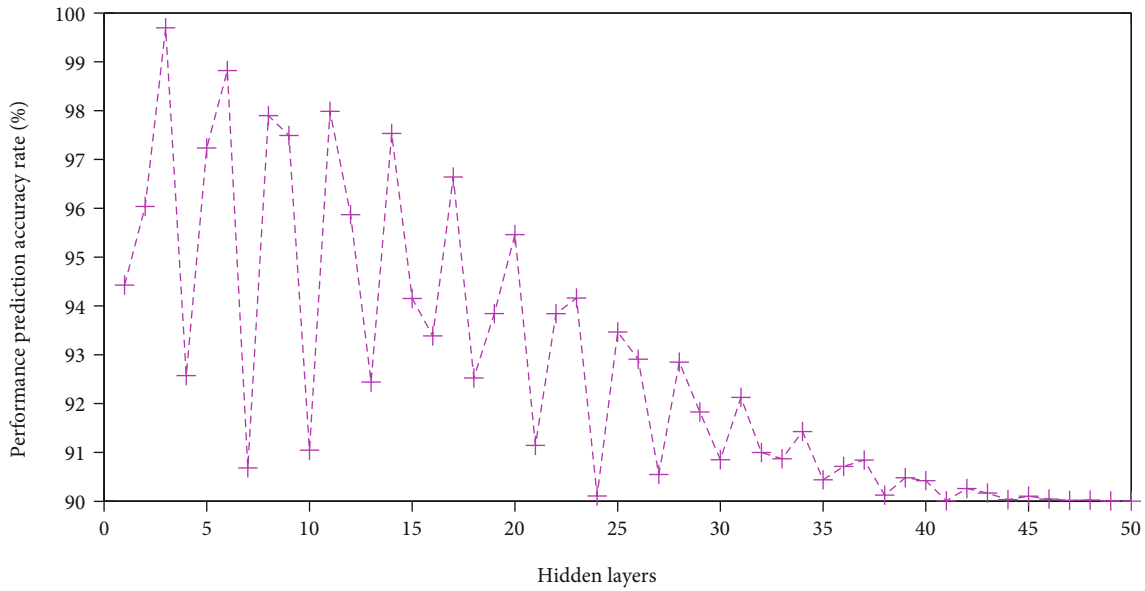


FIGURE 5: The influence of the number of hidden layers in the network on the accuracy of performance prediction.

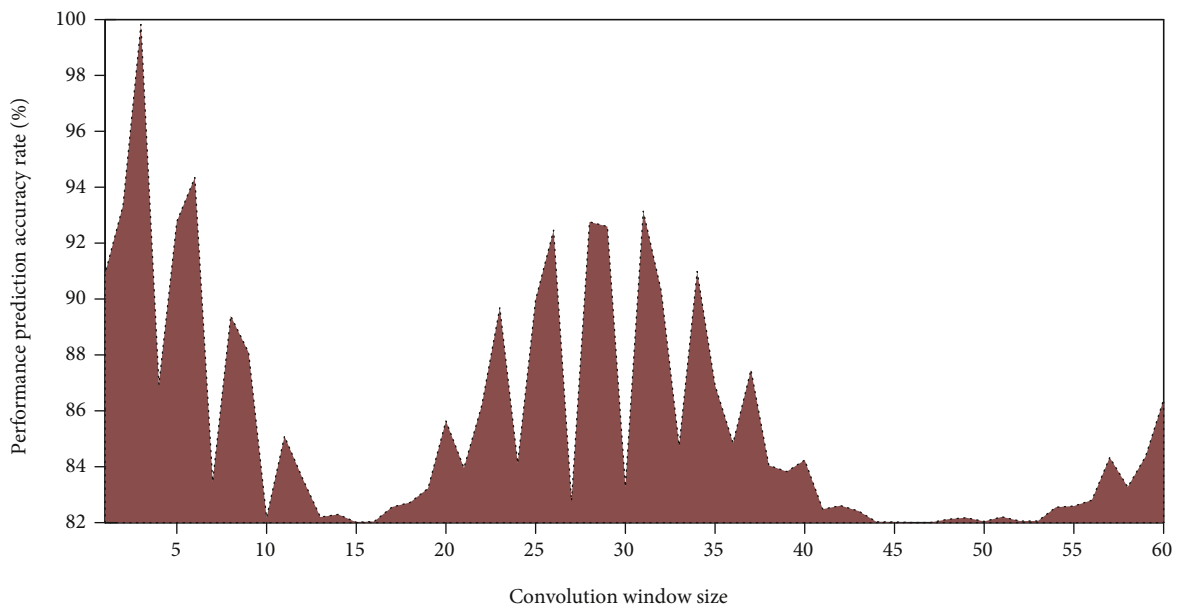


FIGURE 6: The trend graph of the accuracy of performance prediction with the size of the convolution window.

5.4. *Experiments on the Contribution of Different Characteristics to the Performance Prediction of Technological Innovation Enterprises.* I explored the contribution of each type of feature to the performance prediction of technological innovation enterprises, and the experimental results are shown in Figure 7. Removing any type of features will reduce the prediction effect of the experiment. Location features account for 18% of the time, which is in the middle of the contribution of various features.

I also found that whether it is in deep neural networks or convolutional neural networks, the direct use of enterprise difference features will not improve the effect of the experiment, but will have side effects. The reason is that the differ-

ence feature of the enterprise has been processed by the neural network in the extraction, and a vector is used to represent a difference feature of the enterprise, and the other features are the original features in the form of scalar. In the learning process, it does not blend well.

5.5. *Comparative Experiment of Different Models.* I use logistic regression model, support vector regression model, and random forest as comparison methods to compare the prediction effect of the DBN-SVR model. The result is shown in Figure 8.

It can be seen that the DBN-SVR model has the best effect, once again verifying that in the process of feature

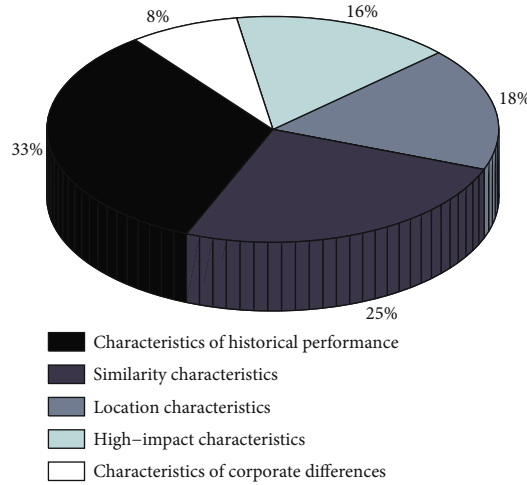


FIGURE 7: Comparison of contribution rates of five types of features.

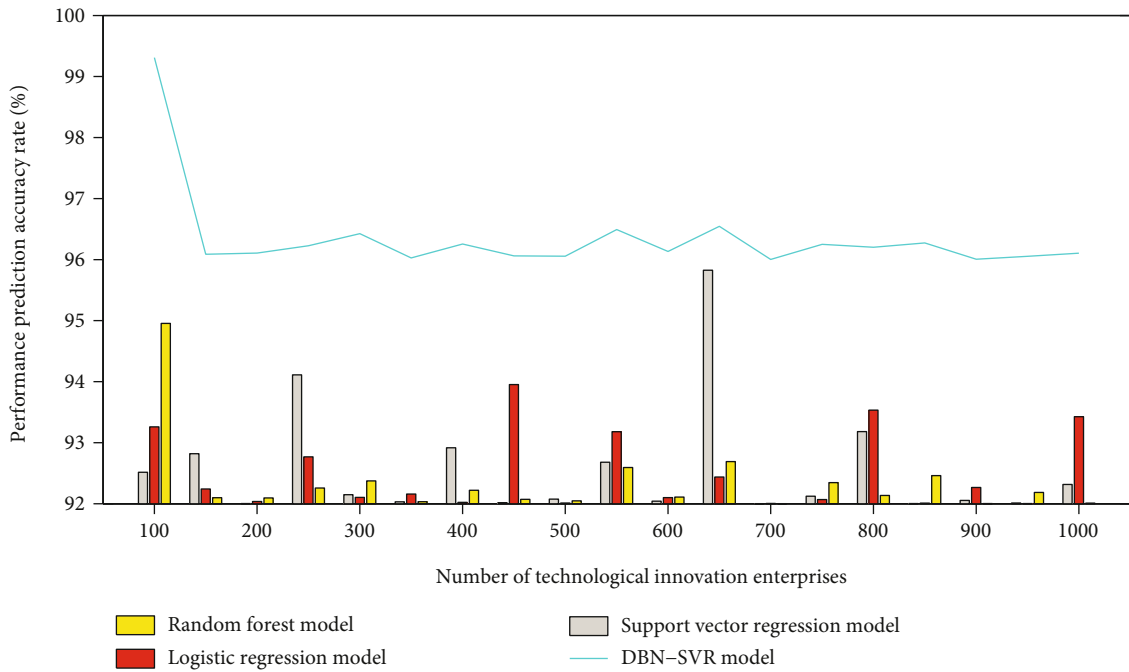


FIGURE 8: Comparison of experimental results of different models.

learning, the deep learning model can better dig the relationship between features than the shallow learning model. This shows that the feature fusion of the convolutional layer and the feature extraction process of the subsampling layer are effective.

### 6. Conclusion

While paying attention to R&D investment, high-tech enterprises should also consider the effect of non-R&D investment on corporate performance. When a company has a variety of different resources to invest together, it should consider how to effectively integrate the resources and use them rationally, optimize the combination of resource inputs, enhance scientific and technological innovation

capabilities, and ultimately achieve the improvement of innovation performance. As some enterprises fail to use resources rationally, leading to the phenomenon of “high input, high consumption, low output, and low added value,” high-tech enterprises should actively try to avoid them. After adjusting and optimizing the investment structure, high-tech enterprises should further improve their technological innovation mechanism, optimize and enhance the quality and efficiency of investment, and effectively boost corporate performance. This paper uses the DBN model to pretrain the large sample data layer by layer and then connects the SVR model to the top of the model to make predictions. In this paper, a large sample data is used for experimental simulation, and the model performance is analyzed by means of mean square error (MSE) and mean absolute percentage

error (MAPE). The influence of the number of nodes in different hidden layers on the prediction error is comparatively analyzed, and multiple prediction models are used to, respectively, predict and compare the prediction effects. Aiming at the problem of slow training time due to the complexity of deep learning models, I designed an experimental method based on GPU computing. The advantages of GPU integration of many computing units make it possible to train deep learning models in the context of big data. In addition, the amount of historical data is very large, and the demand for video memory is high. In the case of limited video memory, I used the idea of block computing to divide the data into blocks for training and testing. Experiments show that within the acceptable range of error, the GPU-based block calculation method can achieve certain experimental results.

## Data Availability

The data used to support the findings of this study are included within the article.

## Conflicts of Interest

The author does not have any possible conflicts of interest.


## References

- [1] H. Jin, X. Chen, L. Wang, K. Yang, and L. Wu, "Dual learning-based online ensemble regression approach for adaptive soft sensor modeling of nonlinear time-varying processes," *Chemometrics and Intelligent Laboratory Systems*, vol. 151, pp. 228–244, 2016.
- [2] H. M. Truong, "Integrating learning styles and adaptive e-learning system: current developments, problems and opportunities," *Computers in Human Behavior*, vol. 55, pp. 1185–1193, 2016.
- [3] X. Yuan, Y. Wang, C. Yang, Z. Ge, Z. Song, and W. Gui, "Weighted linear dynamic system for feature representation and soft sensor application in nonlinear dynamic industrial processes," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 2, pp. 1508–1517, 2018.
- [4] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *International Conference on Machine Learning*, vol. 70, pp. 1126–1135, Sydney, NSW, Australia, 2017.
- [5] X. X. Wang Li and V. C. M. Leung, "Artificial intelligence-based techniques for emerging heterogeneous network: state of the arts opportunities and challenges," *IEEE Access*, vol. 3, pp. 1379–1391, 2015.
- [6] M. Varisco, C. Johnsson, J. Mejvik, M. M. Schiraldi, and L. Zhu, "KPIs for manufacturing operations management: driving the ISO22400 standard towards practical applicability," *IFAC-Papers OnLine*, vol. 51, no. 11, pp. 7–12, 2018.
- [7] Y. Han, Q. Liu, C.-K. Wen, S. Jin, and K.-K. Wong, "FDD massive MIMO based on efficient downlink channel reconstruction," *IEEE Transactions on Communications*, vol. 67, no. 6, pp. 4020–4034, 2019.
- [8] X. Yuan, Z. Ge, B. Huang, and Z. Song, "A probabilistic just-in-time learning framework for soft sensor development with missing data," *IEEE Transactions on Control Systems Technology*, vol. 25, no. 3, pp. 1124–1132, 2017.
- [9] J. Schmidhuber, "Deep learning in neural networks: an overview," *Neural Networks*, vol. 61, pp. 85–117, 2015.
- [10] Y. Yang, F. Gao, X. Ma, and S. Zhang, "Deep learning-based channel estimation for doubly selective fading channels," *IEEE Access*, vol. 7, pp. 36579–36589, 2019.
- [11] T. Vafeiadis, K. I. Diamantaras, G. Sarigiannidis, and K. C. Chatzisavvas, "A comparison of machine learning techniques for customer churn prediction," *Simulation Modelling Practice and Theory*, vol. 55, pp. 1–9, 2015.
- [12] Z. Zhen and Y. Yao, "Optimizing deep learning and neural network to explore enterprise technology innovation model," *Neural Computing and Applications*, vol. 33, no. 2, pp. 755–771, 2021.
- [13] M. L. Borrajo, B. Baruque, E. Corchado, J. Bajo, and J. M. Corchado, "Hybrid neural intelligent system to predict business failure in small-to-medium-size enterprises," *International Journal of Neural Systems*, vol. 21, no. 4, pp. 277–296, 2011.
- [14] A. Abudurehman, A. Nilupaer, and Y. He, "Performance evaluation of enterprises' innovation capacity based on fuzzy system model and convolutional neural network," *Journal of Intelligent & Fuzzy Systems*, pp. 1–9, 2020.
- [15] M. J. Beynon, P. Jones, D. Pickernell, and G. Packham, "Investigating the impact of training influence on employee retention in small and medium enterprises: a regression-type classification and ranking believe simplex analysis on sparse data," *Expert Systems*, vol. 32, no. 1, pp. 141–154, 2015.
- [16] A. Westerski and C. A. Iglesias, "Exploiting structured linked data in enterprise knowledge management systems: an idea management case study," in *2011 IEEE 15th international Enterprise distributed object computing conference workshops*, pp. 395–403, Helsinki, Finland, 2011, August.
- [17] A. M. Pérez, J. M. Gómez, and C. P. Risquet, "Semantic interaction in enterprise data-flow visualization environments: an exploratory study," in *International conference on ICT innovations*, pp. 217–226, Berlin, Heidelberg, 2009, September.
- [18] S. Hassan, H. Waheed, N. R. Aljohani, M. Ali, S. Ventura, and F. Herrera, "Virtual learning environment to predict withdrawal by leveraging deep learning," *International Journal of Intelligence Systems*, vol. 34, no. 8, pp. 1935–1952, 2019.
- [19] F. A. A. Souza and R. Araújo, "Mixture of partial least squares experts and application in prediction settings with multiple operating modes," *Chemometrics and Intelligent Laboratory Systems*, vol. 130, no. 15, pp. 192–202, 2014.
- [20] W. Xing, X. Chen, J. Stein, and M. Marcinkowski, "Temporal predication of dropouts in MOOCs: reaching the low hanging fruit through stacking generalization," *Computers in Human Behavior*, vol. 58, pp. 119–129, 2016.
- [21] H. He, S. Jin, C.-K. Wen, F. Gao, G. Y. Li, and Z. Xu, "Model-driven deep learning for physical-layer communications," *IEEE Wireless Communications*, vol. 26, no. 5, pp. 77–83, 2019.
- [22] C.-K. Wen, W.-T. Shih, and S. Jin, "Deep learning for massive MIMO CSI feedback," *IEEE Wireless Communications Letters*, vol. 7, no. 5, pp. 748–751, 2018.
- [23] J. Zhang, H. Chen, S. Chen, and X. Hong, "An improved mixture of probabilistic PCA for nonlinear data-driven process monitoring," *IEEE transactions on cybernetics*, vol. 49, no. 1, pp. 198–210, 2019.

- [24] J. Zhang, J. Sun, J. Wang, and X. G. Yue, "Visual object tracking based on residual network and cascaded correlation filters," *Journal of ambient intelligence and humanized computing*, pp. 1–14, 2021.
- [25] J. Zhang, X. Jin, J. Sun, J. Wang, and A. K. Sangaiah, "Spatial and semantic convolutional features for robust visual object tracking," *Multimedia Tools and Applications*, vol. 79, no. 21–22, pp. 15095–15115, 2020.
- [26] J. Guo, C.-K. Wen, S. Jin, and G. Y. Li, "Convolutional neural network-based multiple-rate compressive sensing for massive MIMO CSI feedback: design simulation and analysis," *IEEE Transactions on Wireless Communications*, vol. 19, no. 4, pp. 2827–2840, 2020.
- [27] Z. Ge and X. Chen, "Dynamic probabilistic latent variable model for process data modeling and regression application," *IEEE Transactions on Control Systems Technology*, vol. 27, no. 1, pp. 323–331, 2019.
- [28] K. Wang, M. G. Forbes, B. Gopaluni, J. Chen, and Z. Song, "Systematic development of a new variational autoencoder model based on uncertain data for monitoring nonlinear processes," *IEEE Access*, vol. 7, pp. 22554–22565, 2019.
- [29] X. Yuan, B. Huang, Y. Wang, C. Yang, and W. Gui, "Deep learning-based feature representation and its application for soft sensor modeling with variable-wise weighted SAE," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 7, pp. 3225–3243, 2018.
- [30] C. Fan, F. Xiao, and Y. Zhao, "A short-term building cooling load prediction method using deep learning algorithms," *Applied Energy*, vol. 195, pp. 222–233, 2017.

## Research Article

# Research on the Effect of Big Data Flipped Classroom Combined with Scenario Simulation Teaching: Based on Clinical Practice of Medical Students

Shuli Liu,<sup>1</sup> Yong Li,<sup>2</sup> Xiaobo Wang,<sup>1</sup> Xuena Zhang,<sup>1</sup> and Renshu Wang<sup>1</sup> 

<sup>1</sup>Wenzhou Central Hospital ICU, Wenzhou 325000, China

<sup>2</sup>Emergency Department of Wenzhou Central Hospital, Wenzhou 325000, China

Correspondence should be addressed to Renshu Wang; wangrenshu2017@163.com

Received 5 August 2021; Revised 24 August 2021; Accepted 25 August 2021; Published 24 September 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Shuli Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the rapid development of information technology and the deepening of education reform, flipped classroom, as a new teaching mode, has received considerable attention from the education circle as soon as it appeared. This paper analyzes the teaching design principles of big data flipped classroom combined scenario simulation, constructs the teaching design model of big data flipped classroom combined scenario simulation in medical clinical practice, and explores the application and effect of the teaching method of big data flipped classroom combined scenario simulation. In this paper, the experimental group takes care of fusion simulations combined with the flipped classroom teaching method, and the routine control group uses the traditional teaching method; the scores of the two groups on theory, learning initiative, clinical practice, and critical thinking ability allow comparing the teaching effect, concluding that health care integration scenario simulation combined with flipped classroom teaching methods is new. It is beneficial to improve the learning ability, improve the ability of medical cooperation, enhance the sense of humanistic care, and so on. Analysis and comparison concludes that turning medical fusion simulations combined with classroom teaching, promoting the medical students on medical students' clinical practice to master theoretical knowledge and practical operation ability, and improving the ability of critical thinking ability and care will lead medical students to make a study of basic knowledge of medicine helping to cultivate the graduate student's study enthusiasm, stimulating interest in learning, improving the comprehensive quality of medical specialty, and making it worth popularizing in practice teaching.

## 1. Introduction

With the progress of science and technology and the development of education, medical teaching in colleges and universities should not only adapt to the needs of social development in the new period but also introduce and fully implement the new teaching mode in the teaching concept classroom [1] in time, so as to improve the classroom teaching effect and cultivate students' comprehensive ability. On the teaching platform, the learning situation is statistically sorted, and the focus, time, habit, and effect of students' learning are analyzed online. From the educational field, students are provided with quality, personalized learning experience. In the context of the era of deep learning [2–4], learning analysis

can be used to collect and analyze data for each link in the teaching process, so as to use a more appropriate teaching mode to improve the teaching effect, improve students' learning environment, and improve their learning enthusiasm. With the deepening of research, the concept of big data is generally regarded as a huge data set; through the collection, processing, and analysis of the data set, valuable information is obtained. The big data involved mainly refers to information, data, and technology in the field of education. It refers to the rapid and efficient analysis and processing of the learning data through the collection of a large number of learners' learning data. It is the internal unification of the technology of big data and the application in the field of education [5], and the concept and means to realize human

development and the improvement of education and teaching with the ultimate goal of educating people [6]. With the support of the continuous development and progress of information network technology, the data information is analyzed and processed in scientific mode, and the value of data is found in the process of analysis. Big data technology in education provides reliable basis for education and teaching through data analysis. Current researches mainly focus on the theoretical discussion of big data in the field of education. There is still a lack of research on the practical experience and application effect of big data in teaching, and there is no more systematic research. Therefore, it is still very necessary and urgent to further develop the breadth and depth of research on big data in teaching practice, so that big data can play its unique role in the field of education as soon as possible.

Scenario simulation teaching method, as a virtual practical training method [7], has become an increasingly popular teaching method in the field of international education. This method can provide students with a safe, nonthreatening, efficient and realistic simulation environment [8]. Application in clinical practice is to make full use of the intuitiveness, practicability, effectiveness, repeatability, and control of drills [9]. Promoting clinical situational teaching can enable students to develop their abilities of observation, understanding, analysis, and problem solving in real cases and situations, and at the same time exercise and embody students' comprehensive qualities such as cooperation spirit, humanism, and critical thinking [10]. Bates and Ludwig pointed out that the clinical teaching method combined with the PBL model and scenario simulation cultivated the trainees' critical thinking ability, improved the trainees' comprehensive skills of emergency surgery cooperation, and enhanced the trainees' team communication ability and cooperation spirit [11]. Chen et al. pointed out that the situational simulation teaching method of medical record introduction improved the language communication and necessary change ability of nursing students, and enhanced their listening skills [12]. Choi et al. applied the scenario simulation teaching method of medical record introduction to students' teaching, and believed that the method improved students' theoretical exam results, strengthened students' teamwork spirit, and improved students' learning enthusiasm and teaching satisfaction with teachers [13].

Students are actively involved in learning, and personalized learning is designed for students. As a new teaching mode, the flipped classroom replans the roles of teachers and students in traditional teaching, which is an innovation of traditional teaching. Hanson combs and classifies the flipped classroom network AIDS that is popular among teachers and students, and makes a detailed introduction to AIDS [14]. Based on STEM concepts, Sandrone et al. constructed the MOOC-platform-based flipped classroom teaching mode and applied it to the university computer foundation, achieving good results [15]. Fahy et al. drew on the teaching model designed by a network platform to provide convenience and guidance for teachers in the teaching practice, studied the implementation of the model based

on the information platform, and put forward constructive suggestions [16]. Saunders et al. pointed out that after the flipped mode, students' attention became more focused, the classroom atmosphere became more flexible and interesting, and students could actively participate in the class in cooperative learning [17]. The existing research on big data in education is still only a "point" exploration, lacking a deep and comprehensive discussion on the development of education by big data, and the research on the practical application of big data in the field of education is also insufficient. Combined with the flipped classroom teaching mode, many problems are found in the implementation process. In this paper, the advantages of big data technology and the flipped classroom scenario simulation teaching mode are combined to explore the practical application of big data technology in flipped classroom teaching and new methods [18] to solve the flipped classroom teaching mode.

## 2. Big Data Flipped Classroom Combined with Situational Simulation Teaching

*2.1. Flipped Classroom Combined with Situational Simulation.* Flipped classroom refers to the reversal of the roles of the student and the teacher. The traditional classroom is mainly based on the teacher as the main body, through the teacher as the master, ignoring the students' subjective initiative and resulting in the lack of students' subjective initiative. It is difficult to have their own independent creative thinking, and such a teaching model not only weakens the students' interest in learning but also limits the development of students' innovative thinking and makes students' minds become rigid and become "robots" that can only accept passively but cannot innovate independently. However, if we turn in the classroom the study of sovereignty back to the students, students can grasp the knowledge in its heaviest and most difficult part, according to the needs of individuals for the consolidation and repetition of the target, while improving the learning efficiency to reduce many of the useless things done by the teacher in class, to a certain extent, reducing the financial burden on the teacher's teaching. On the other hand, a flipped classroom also means readjusting the time inside and outside the classroom and breaking the original time organization mode of the classroom. The comparison with a traditional classroom is shown in Figure 1:

Situational simulation teaching mainly refers to a method where the teacher imparts the teaching content according to the needs and sets up relevant teaching situations in advance in combination with the actual situation, and then guides the students to play different roles in the specific scene, so as to simulate the actual problems and solve them [19]. The combined teaching mode of FC and SST can fully arouse students' learning enthusiasm and initiative, and have more practical opportunities to improve students' ability to solve practical problems and improve the teaching effect. Flip joint scene simulation teaching is to turn class and class scene simulation teaching which is a combined application of a kind of teaching method. The teaching method is mainly realized with the help of



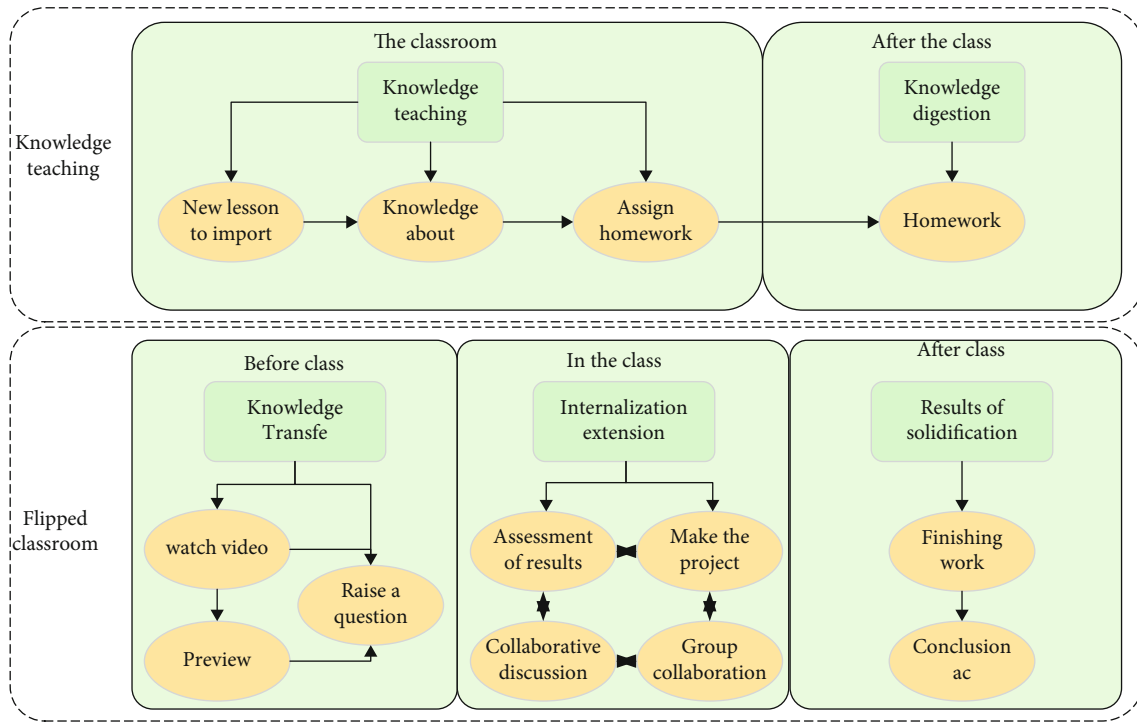


FIGURE 1: Flipped classroom versus traditional classroom.

multimedia technology, which requires teaching videos, teaching construction, and uses scenarios to guide students to learn accordingly. This can not only improve the interest of teaching activities but also promote the comprehensive quality of students [20]. Moreover, this teaching method is widely used in public health, teratology, clinical medicine, nursing, and pharmacy, which plays a very positive role in the realization of the goals of modern medical education in China.

**2.2. Theoretical Basis of Flipped Classroom Combined with Scenario Simulation Teaching.** From the perspective of a flipped classroom, constructivism education emphasizes the initiative and constructivism of students, the importance of group cooperative learning and situational learning, and the practical application of educational information technology in teaching [21]. The implementation of a flipped classroom is supported by mastering learning theory and emphasizes on facing all students to meet the learning needs of students. When making learning objectives, we should pay attention to the individual differences of students, choose different learning materials, and adopt different teaching methods for different students. The learning process focuses on students' emotional experience, so that students can get a sense of achievement and enjoy the happiness of learning, and develop students' self-concept deeply. The course promotes the cooperation between students as well as the communication between teachers and students and the cultivation of cooperative spirit.

Development zone refers to the distance between a student's actual level of development when solving problems independently and the potential level of development when

solving problems with teachers and partners. Therefore, the first level of development is the level of development that students can achieve through independent learning with their own abilities and without external help. The second level of development is the level of development that students cannot achieve by themselves with their own abilities, but can only be completed through cooperative learning with the help of teachers or partners. Therefore, the key of teaching design for teachers is to identify the nearest development area for students. On this basis, teachers intervene in students' learning and build a series of "scaffolding" for students' self-construction knowledge.

Flipped classroom focuses on the individualized development of students, and pays attention to the new view of individualized teaching and hierarchical teaching based on the theory of proximal development zone. While the classroom teaching mode of teaching design focuses on the discovery of students, the zone of proximal development focuses on the student's existing level; this is not only reflected in the setting of teaching content but also reflected in teaching in the teaching video speed where each student can make their own decision according to their own situation. It is repeatable, either for the students' personalized learning, or for excavating its potential and innovation ability. It allows students to move beyond the nearest development zone to a higher level. Based on the learning of this theory, the structure diagram of the flipped classroom combined scenario simulation teaching relationship based on big data is given, as shown in Figure 2:

**2.3. Technical Characteristics of Big Data.** To combine big data technology with education and teaching well, it is

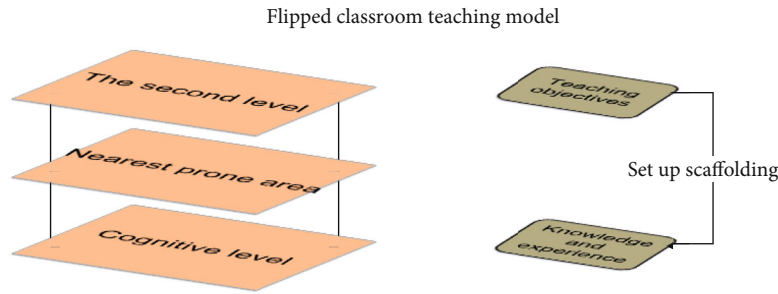


FIGURE 2: Big data flipped classroom combined with situational simulation teaching relationship structure diagram.

necessary to understand the characteristics of big data technology so as to give full play to its strengths and advantages [22]. Big data technology can process a huge amount of data. In the era of big data, the data in people's life and study will grow every day. The use of big data technology can solve the accumulation of a large amount of data, such as text data, image data, and learning process data. Through various wired and wireless networks, communication between information subjects can take place anytime and anywhere, which inevitably brings information exchange. The key to the effectiveness of data and information lies in the improvement of speed, and big data technology can process data efficiently and quickly. The application of big data technology has the effect of low density and high value. Different types of data use big data analysis technology to mine out the hidden value in information, so as to facilitate study and research, and realize the convenience and depth of data mining.

The components of big data mainly include the process data of students' learning, including the viewing of teaching videos and teaching PPT and the completion of task lists. After the data is analyzed by the big data technology of the cloud platform, teachers will look for the starting point of teaching and create an appropriate teaching situation based on the data analysis result feedback from the cloud platform. This teaching situation can sort out, review, and consolidate the knowledge that students have mastered, and at the same time, introduce the knowledge content and ability points that most students do not know, so as to carry out classroom teaching, so as to realize the teaching of a flipped classroom under the background of big data.

### 3. Design and Implementation of Flipped Classroom Combined with Situational Simulation Teaching

*3.1. Teaching Design Principles of Big Data Flipped Classroom Combined with Scenario Simulation.* One of the basic jobs of a teacher is to prepare lessons before they begin. In order to have a good class, it is necessary to carry out reasonable and effective teaching design in advance, rather than simply reading the textbook. Flipped classroom is an emerging teaching mode. With the support of information technology, flipped classroom reverses the order of transmission and internalization of learning content and breaks the teaching mode of a traditional classroom. It is

necessary for teachers to replan the teaching design process of a flipped classroom. In the combined scenario simulation teaching design based on a big data flipped classroom, the roles of teachers and students have changed. Teachers are no longer the porter of knowledge, but the guide, giving play to the leading and organizational role of teachers. Students are the main body of classroom learning; the classroom should be returned to the students, transforming students from becoming passive to being active. Classroom learning is a bilateral activity between teachers and students, with teachers guiding and students actively participating. In class, the teacher should grasp the question before the students, and they can be understood by some students and allow them to speak their views, can encourage good students, and can stimulate the underachievers. This allows the students, in cooperation with teachers and peers, to complete the task process in-depth knowledge.

Classroom interaction communication refers to the process in which teachers use verbal or nonverbal information to stimulate one or more students' thoughts in class, so that teachers and students interact with each other. In turn classroom teaching, teachers should first have a deeper understanding to the essence of classroom interaction and accurately grasp the question, which cannot simply be equated to communication, but should be based on students' feedback learning before class, to understand their knowledge before class, put forward reasonable effective study, where students explore cooperation, communication between teachers and students during discussion, and personalized guidance. At the same time, the more communication, the better; teachers should control the overall situation and grasp the degree of classroom interaction, so as to ensure the effectiveness of interaction and communication.

*3.2. Construction of Teaching Design Model of Flipped Classroom Combined with Scenario Simulation Based on Big Data.* The contents of flipped classroom teaching design mainly include preclass knowledge acquisition and classroom knowledge internalization. At the preclass stage, students study and discuss the course content in advance. According to the essence of a flipped classroom, a combined scenario simulation teaching design based on big data flipped classroom is set. The visual influence of video and sound of the teaching contents is recorded, which repeatedly impact the visual memory of students, alleviate the thinking fatigue caused by students from simply reading books, and

deepen the firmness of knowledge storage in the brain. It provides a basis for students to further study theoretical knowledge after class, and can timely and accurately “answer questions” for students, timely correct students’ wrong views, and improve the accuracy of knowledge, as shown in Figure 3:

Normalize flipped classroom teaching model under big data thinking. Horizontally, the model includes three parts: before class, during class, and after class. Vertically, the model includes two parts: the normal flipped classroom teaching process and the learning analysis system. The data collected before class is the basis for effective classroom teaching, while the data collected during and after class is not only the basis for adjusting teaching pace and carrying out personalized guidance but also the evidence for teaching students in accordance with their aptitude and promoting hierarchical teaching.

Flip the classroom scene simulation teaching has related advantages in application of modern information technology; in network platform for related animation, film, video, and picture upload, the related knowledge points through multimedia are intuitively shown in front of the students, which improves students’ learning interest and participation in teaching, to help them better understand relevant knowledge. In clinical medicine teaching, students master the corresponding medical knowledge and have the corresponding knowledge application ability. In the teaching design, classroom organization form diversified designs, in order to create scenes, for example, using a computer to control patient models, and simulating signs and symptoms of the patients with students in the class to learn the knowledge, simulating the real clinical practice in the classroom, and carrying on the application of relevant knowledge, which enable students to further internalize the knowledge.

**3.3. Teaching Research Tool.** The active learning ability table mainly evaluates students’ initiative in learning for medical and related majors [23]. There are 21 items in the scale of active learning ability, including learning driving force, learning goal, deep learning, control learning, and solid learning. The expert content validity index and the consistency of the raters of the scale were both 0.861, and the overall consistency reliability coefficient A was 0.895. The learning initiative scale was divided into multiple dimensions, with 1-5 points for each question.

The Mini Clinical Exercise Evaluation Scale is a set of tools [24] developed on the basis of the traditional evaluation scale to evaluate and feedback clinical knowledge and skills. It has been used as a clinical teaching method for residents or medical students. The content validity index and the consistency of the raters of the scale were both 0.8675, and the overall consistency reliability coefficient A was 0.8933. The scale had a total of seven dimensions and was scored on a 9-point scale with 3 levels. Scores with 1-3 points are judged to be strengthened, scores with 4-6 points are judged to be in line with the requirements, and scores with 7-9 points are judged to be excellent.

The critical thinking trait scale has seven dimensions [25], and each dimension has 10 items, for a total of 70

items, with 30 positive items and 40 negative items. Each item was rated on a scale ranging from “strongly agree” to “strongly disagree,” with positive items being rated 6-1 and negative items being rated backwards. The scale had a maximum score of 420. A score > 350 indicates that the subjects have a strong trait of critical thinking, >280 indicates that the subjects have positive critical thinking, <280 indicates that the critical thinking is weak, and <260 indicates a lack of critical thinking skills. A score of less than 30 on each subscale indicates negative trait performance, a score of 31-39 on the subscale indicates moderate thinking ability, and a score of 40-49 on the subscale indicates positive trait performance. According to the research, the A value of CTDI-CV is 0.90, and the A value of trait is 0.54 to 0.77, which shows a high internal consistency and can more accurately reflect students’ critical thinking ability [26].

**3.4. Preparations before Implementation.** According to the clinical teaching syllabus of paediatrics in the hospital, classic clinical diseases of paediatric neurology were selected to explain children’s growth and development, paediatric common auxiliary examination, child convulsions, epilepsy, and paediatric viral encephalitis. Each disease was explained by the same teacher, and the class time was controlled within 90 minutes. According to the clinical practice of hospitals for clinical nursing practice teaching outline and the actual situation, the department wrote four scenarios for a simulation script: new hospital information collection of children with viral encephalitis, sedative drug retention enema for children with operations, families of children with high fever given measures of health education, and emergency treatment of children with convulsions. According to the content, each script has one child, one child’s family member, and one nurse. The content and lines of the script are made by the teacher through discussion, and then implemented after having been reviewed and modified by two neurologist nursing experts in our hospital. The role of the nurse is played by the teacher, and the role of the child and the role of the family member is recommended by the group. Each scenario simulation class is taught by the same teacher, and she has two teachers to help. The teaching teacher is a researcher, and there are two teachers from the department. Both of them have a bachelor’s degree and have the title of supervisor nurse. All of them have passed the hospital certification. Before the official classes taught by the teacher for two groups of students, the purpose is to ensure the smooth progress of the experiment, mainly aimed at the scene simulation teaching and the related content, matters needing attention in the video feedback teaching, and how to solve the incident for explanation, knowledge related to organize students to role play group. Before the training, we should follow the teacher’s arrangement, fully realize the importance and necessity of this study, and memorize the steps of the situation simulation and video feedback teaching. Nursing students should read the script carefully and be familiar with the relevant operation steps. Students who will role play should memorize the relevant lines and figure out the role psychology. The operation of nursing skills can be performed on the role of interpretation, and for operations

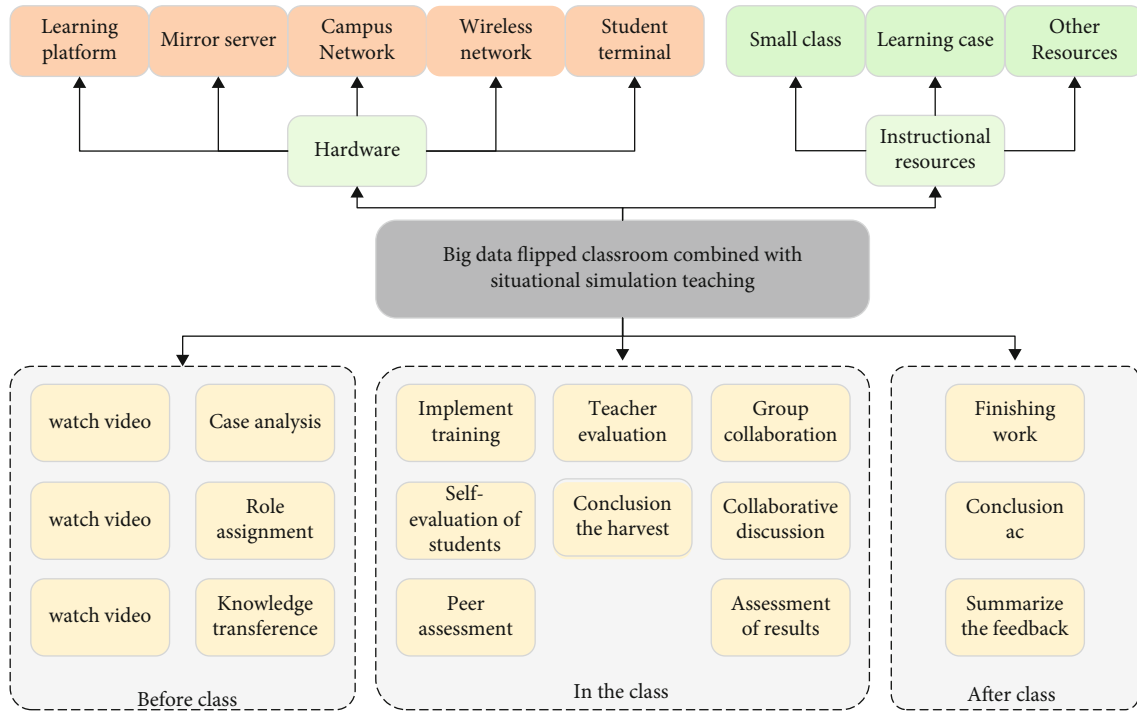


FIGURE 3: Big data flipped classroom combined scenario simulation teaching design model.

TABLE 1: Comparison of the basic situation between the observation group and the control group.

Project	Number of people	Gender		Age	Performance
		Men	Women		
Observation group	61	55	6	22.83 ± 1.235	77.45 ± 7.21
Control group	58	53	5	22.94 ± 1.345	78.48 ± 6.72
	<i>T</i>		—	1.321	1.007
	$\chi^2$		0.053	—	—
	<i>P</i>		0.801	0.179	0.423

that cannot be performed in related cases, such as enema, they can be practiced according to the nursing model.

### 3.5. Implementation Steps of Big Data Flipped Classroom Combined with Situational Simulation Teaching

3.5.1. *Step 1: Script Walkthrough (60 minutes)*. First, the teacher and the role-playing students will perform the simulation of the relevant situation. During the performance, the teacher will explain the relevant theoretical knowledge, operation techniques and key points, communication skills, and matters needing attention while performing. Then, students were asked to play roles in assigned groups. The teacher gave guidance at the right time to guide students to think independently, solve problems in time during the exercise, set obstacles according to the level of the exercise, and record students' cooperation ability when encountering obstacles.

3.5.2. *Step 2: Evaluate Each Other (20 minutes)*. At the end of the exercise, students' self-evaluation was carried out first. Each group member had to speak and tell about their feel-

ings for each role during the exercise, find out their own deficiencies, and put forward improvement methods. During this time, other members of the group recorded and summarized. Secondly, students' mutual evaluation, which includes both evaluation within groups and evaluation between groups, points out the advantages and disadvantages of other students to achieve resource sharing. The teacher leads in asking questions and guides the students to think more deeply. For the problems that students cannot solve by themselves, they are left as homework after the class. Students are allowed to consult related materials and consult related experts, so as to solve the problems and deepen their memory.

3.5.3. *Step 3: Summary of Teacher Evaluation (10 minutes)*. The students will be evaluated in the process of different role rehearsals and discussions, and the students' own performance ability will be evaluated. The teacher summarizes one student's excellent performance and instructs other students to imitate it. In view of the deficiency of students, the

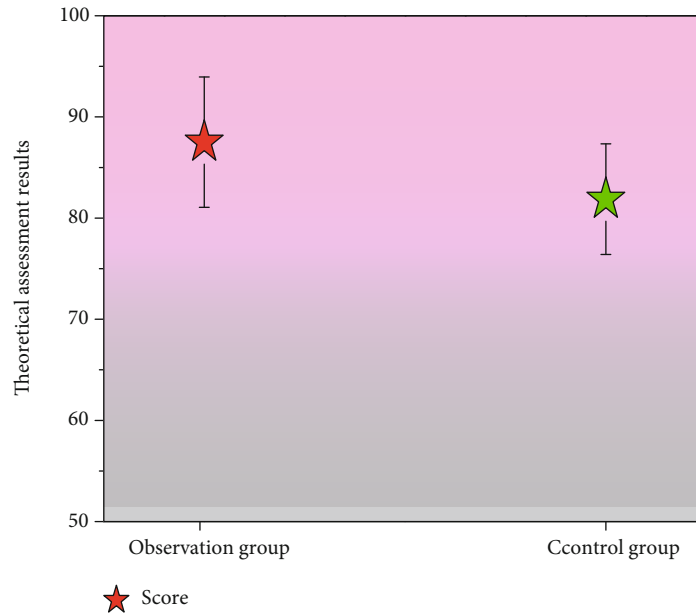


FIGURE 4: Comparison of theoretical assessment results between the observation group and the control group.

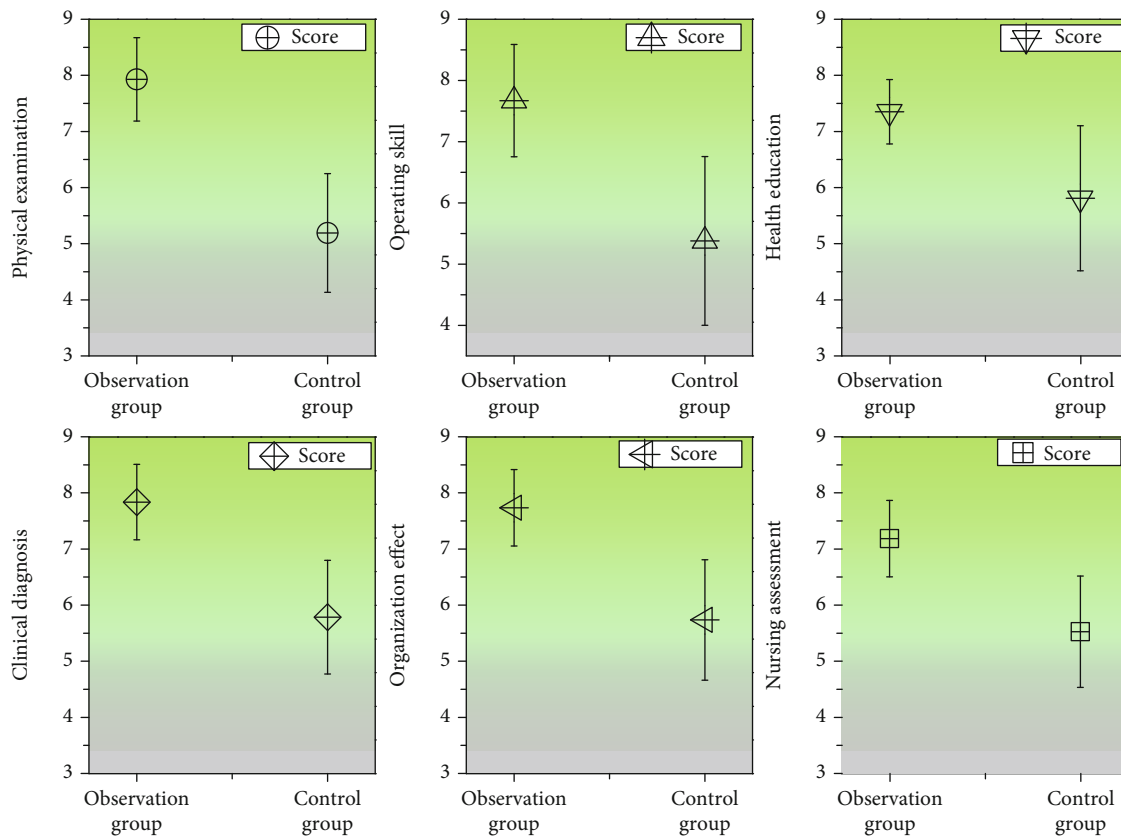


FIGURE 5: Comparison of clinical scores between the observation group and the control group.

teacher asked questions in time to activate students' thinking.

The entire teaching process is videotaped and the students copy the video after class to ensure that students can

watch the teaching video anytime and anywhere. After class, students are required to refer to the teaching videos for operational exercises and theoretical learning, to review and analyze the videos, and to find out their own mistakes and

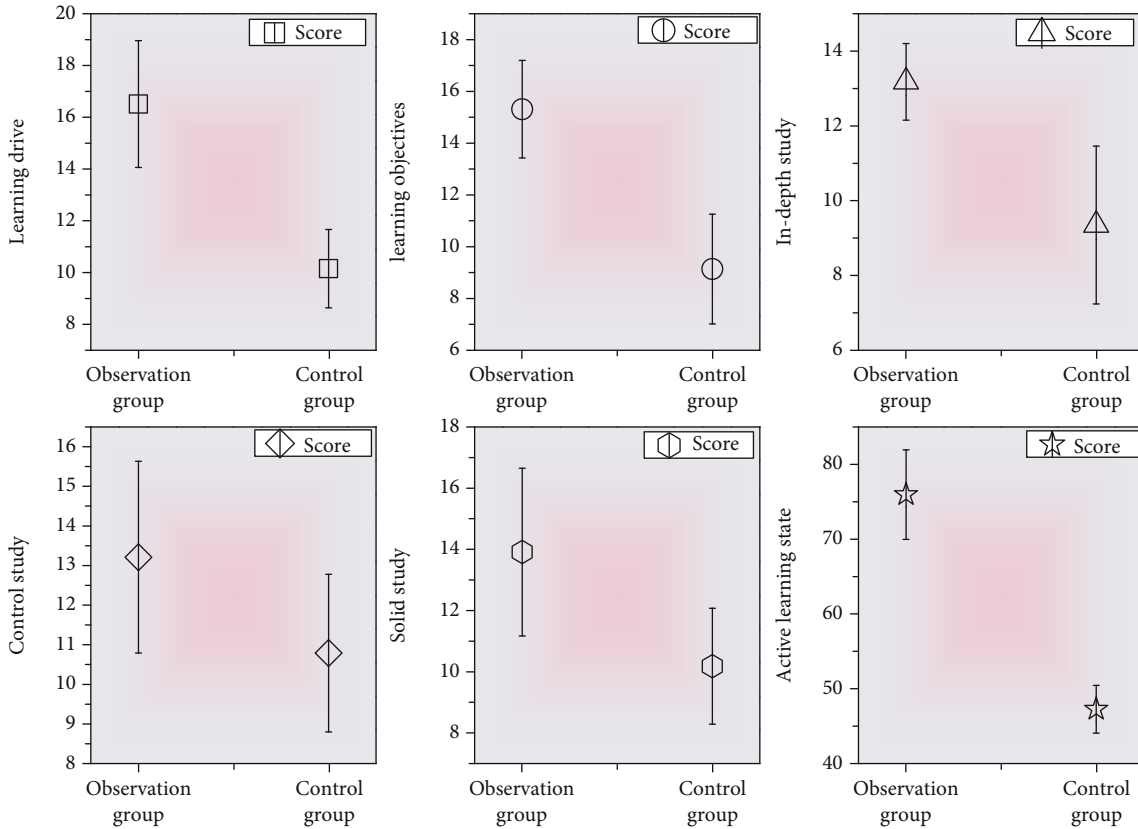


FIGURE 6: Comparison of learning initiative score between the observation group and the control group.

deficiencies through the videos. After class, students can consult the teachers at any time, and hand in written records and exercise videos every week. In the control group, the students only took the abovementioned situation simulation teaching but did not videotape the teaching process. They could consult the teacher at any time after class.

#### 4. Flipped Classroom Combined with Situational Simulation of Teaching Results

**4.1. The Basic Data of Two Groups of Students Were Compared.** Descriptive statistics and an independent sample test were performed on the three aspects of age and theoretical achievement when the two groups were admitted to the hospital. Chi-square test was performed on gender, and the results showed that there was no statistical difference between the two groups. The statistical results of age, sex, and achievement are shown in Table 1, where  $T$  is a statistic in mathematical statistics,  $\chi^2$  is the set sample statistic, and  $P$  value is probability, reflecting the probability of an event happening.

**4.2. Theoretical Score Comparison.** The results of descriptive statistics and an independent sample test showed that the scores of the observation group were higher than those of the control group, indicating that there was a significant difference between the two groups in the assessment of theoret-

ical scores, and the difference was statistically significant. The statistical analysis results are shown in Figure 4.

**4.3. Comparison of Clinical Practice.** The scores of the observation group and the control group in the Mini Clinical Exercise Evaluation Scale were tested by descriptive statistics and an independent sample test, and the results showed that in the main seven assessment aspects of the scale, namely, nursing evaluation, physical examination, operation skills, health education, clinical judgment, organizational effectiveness, and humanistic care, the observation group was better than the control group, and the difference of data was significant, with statistical significance. The specific score is shown in Figure 5.

**4.4. Learning Active Initiative Comparison.** Descriptive statistics and an independent sample test were conducted on the scores of the two groups, and the results showed that the observation group was higher than the control group in various ability indicators, and there was a significant difference between the two groups. See Figure 6 for details.

**4.5. Comparison of Critical Thinking Ability.** After training, the two groups of students were compared on critical thinking ability, and the results showed that the two groups of students' total score of critical thinking ability was  $>280$ ; however, the observation group's total score of critical thinking ability was significantly higher than that of the control

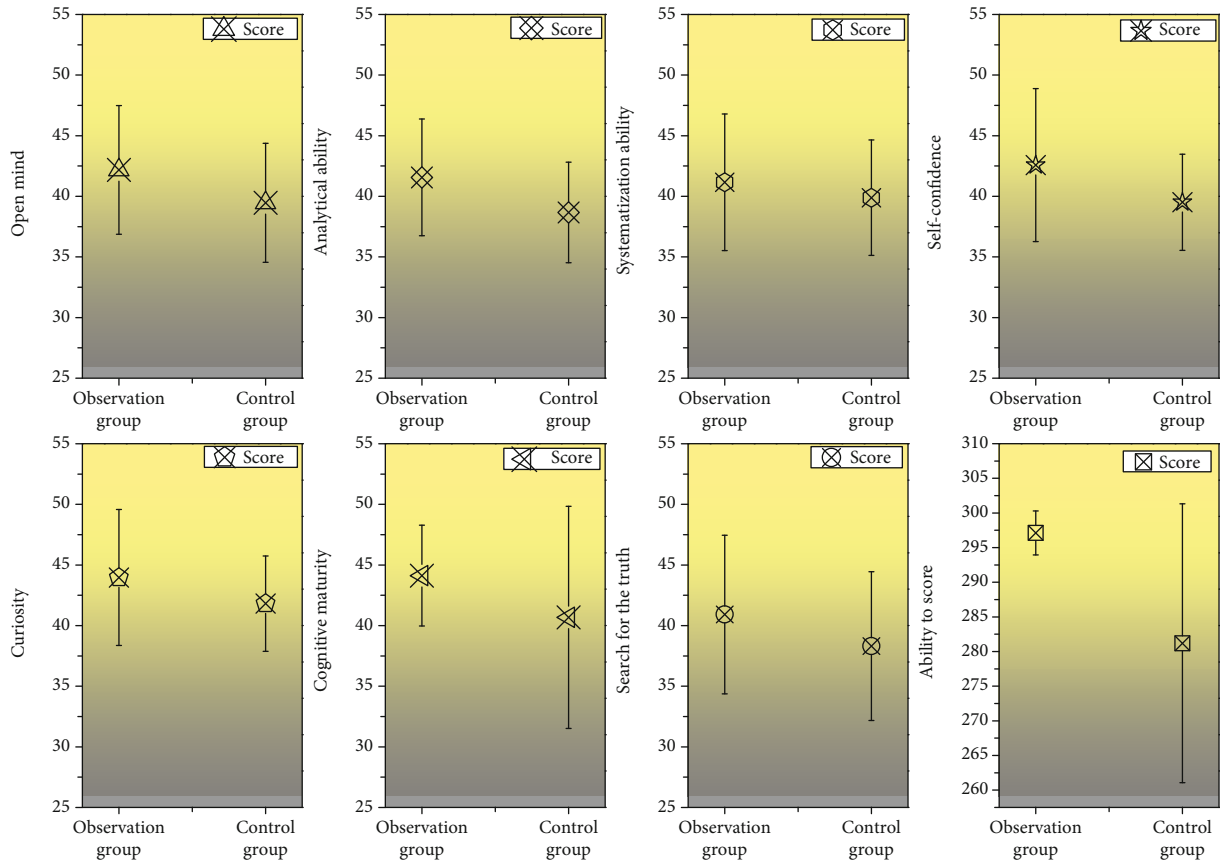


FIGURE 7: Comparison of thinking ability between the observation group and the control group.

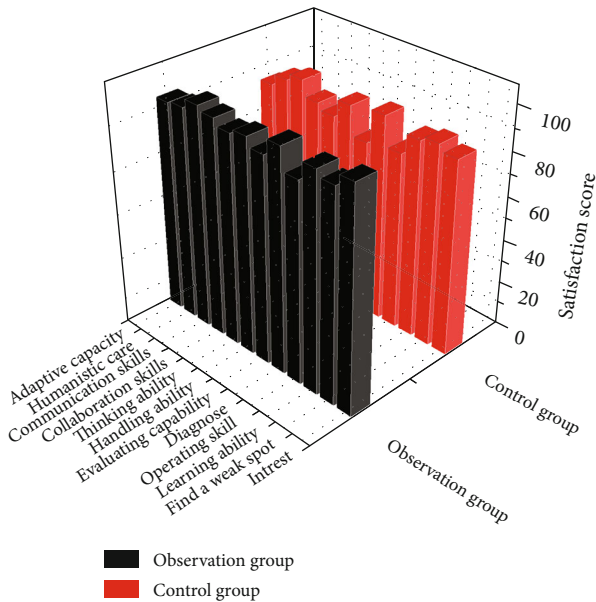


FIGURE 8: Teaching model satisfaction comparison chart.

group, and the difference was statistically significant. There was no statistically significant difference in truth seeking between the two groups, but there were statistically significant differences in open mind, analytical ability, systematic

ability, self-confidence, curiosity, and cognitive maturity between the two groups. The specific score is shown in Figure 7.

4.6. *Teaching Mode Effect Comparison.* On 18 aspects of teaching satisfaction statistics, according to the results by two methods of teaching, there was a certain effect; that is, it shows from comparison that the observation group is significantly higher than the control group. For the observation group of students, teaching methods satisfaction was from 90.2% to 100%, and the control degree of satisfaction was 77.6%-93.1%, from two groups of data tested. The test results showed that there was no statistical significance in the four aspects of timely discovery of knowledge weaknesses, improvement of literature retrieval ability, improvement of team cooperation ability, and improvement of professional identity, while the other 14 differences were statistically significant. Figure 8 shows the proportion of people who showed satisfaction.

## 5. Conclusion

Scenario simulation teaching can combine theoretical knowledge with practice, and it can apply theoretical knowledge to future practice while carrying out theoretical teaching. The results of this study show that, through the flipped classroom teaching mode, students in the

experimental group scored significantly better than the control group in the theoretical assessment, indicating that flipped teaching mode can more effectively strengthen students' mastery of theoretical knowledge. In addition, flipped classroom teaching mode is also conducive to the realization of the close connection between theory and practice, as well as between the major and clinical practice. It helps to train medical students to master the basic theoretical knowledge of diseases in their major, and improve their clinical thinking ability, practical operation ability, and clinical work ability. In this study, the experimental group students' clinical skills examination score is significantly higher than that of the control group, and the prompt turn classroom teaching model can guide students to actively participate in clinical practice skills of learning and training of operation. It is associated with many flip classroom teaching patterns where the results are consistent, and the turn classroom teaching model can improve medical students' knowledge of basic theory and clinical practice ability. It can obviously improve the teaching quality of clinical teaching for medical students. The disadvantages of this model should be overcome in future work: western cultural differences cause the students' deep-rooted educational concept difficult to change, scenario simulation teaching has high requirement on educational resources, and there is lack of an excellent test evaluation system for situational simulation teaching.

### Data Availability

The data used to support the findings of this study are included within the article.

### Conflicts of Interest

All the authors do not have any possible conflicts of interest.

### References

- [1] N. Gostelow, J. Barber, F. Gishen, and A. Berlin, "Flipping social determinants on its head: medical student perspectives on the flipped classroom and simulated patients to teach social determinants of health," *Medical Teacher*, vol. 40, no. 7, pp. 728–735, 2018.
- [2] S. Padugupati, K. P. Joshi, T. V. Chacko, and D. Jamadar, "Designing flipped classroom using Kemp's instructional model to enhance deep learning and self-directed collaborative learning of basic science concepts," *Journal of Education and Health Promotion*, vol. 10, no. 1, p. 187, 2021.
- [3] B. Danker, "Using flipped classroom approach to explore deep learning in large classrooms," *IAFOR Journal of Education*, vol. 3, no. 1, pp. 171–186, 2015.
- [4] I. Le Roux and L. Nagel, "Seeking the best blend for deep learning in a flipped classroom—viewing student perceptions through the Community of Inquiry lens," *International Journal of Educational Technology in Higher Education*, vol. 15, no. 1, pp. 1–28, 2018.
- [5] L. Liang, Q. Yin, and C. Shi, "Exploring proper names online and its application in English teaching in university," *ASP Transactions on Computers*, vol. 1, no. 1, pp. 24–29, 2021.
- [6] P. Xu, Y. Chen, W. Nie et al., "The effectiveness of a flipped classroom on the development of Chinese nursing students' skill competence: a systematic review and meta-analysis," *Nurse Education Today*, vol. 80, pp. 67–77, 2019.
- [7] V. P. S. Njie-Carr, E. Ludeman, M. C. Lee, D. Dordunoo, N. M. Trocky, and L. S. Jenkins, "An integrative review of flipped classroom teaching models in nursing education," *Journal of Professional Nursing*, vol. 33, no. 2, pp. 133–144, 2017.
- [8] T. J. A. Busebaia and B. John, "Can flipped classroom enhance class engagement and academic performance among undergraduate pediatric nursing students? A mixed-methods study," *Research and Practice in Technology Enhanced Learning*, vol. 15, no. 1, pp. 1–16, 2020.
- [9] P. Uther, K. A. van Munster, N. Briggs, S. O'Neill, and S. Kennedy, "Introducing early-phase medical students to clinical paediatrics using simulation and a flipped-classroom," *Journal of Paediatrics and Child Health*, vol. 55, no. 9, pp. 1107–1112, 2019.
- [10] C. R. Presti, "The flipped learning approach in nursing education: a literature review," *Journal of Nursing Education*, vol. 55, no. 5, pp. 252–257, 2016.
- [11] D. Bates and G. Ludwig, "Flipped classroom in a therapeutic modality course: students' perspective," *Research and Practice in Technology Enhanced Learning*, vol. 15, no. 1, pp. 1–15, 2020.
- [12] F. Chen, A. M. Lui, and S. M. Martinelli, "A systematic review of the effectiveness of flipped classrooms in medical education," *Medical Education*, vol. 51, no. 6, pp. 585–597, 2017.
- [13] J. Y. Choi, S. E. Lee, J. Bae et al., "Undergraduate nursing students' experience of learning respiratory system assessment using flipped classroom: a mixed methods study," *Nurse Education Today*, vol. 98, p. 104664, 2021.
- [14] J. Hanson, "Surveying the experiences and perceptions of undergraduate nursing students of a flipped classroom approach to increase understanding of drug science and its application to clinical practice," *Nurse Education in Practice*, vol. 16, no. 1, pp. 79–85, 2016.
- [15] S. Sandrone, J. V. Berthaud, C. Carlson et al., "Strategic considerations for applying the flipped classroom to neurology education," *Annals of Neurology*, vol. 87, no. 1, pp. 4–9, 2020.
- [16] B. G. Fahy, T. Vasilopoulos, and D. F. Chau, "Use of flipped classroom and screen-based simulation for interdisciplinary critical care fellow teaching of electroencephalogram interpretation," *Neurocritical Care*, vol. 33, no. 1, pp. 298–302, 2020.
- [17] A. Saunders, R. Green, and M. Cross, "Making the most of person-centred education by integrating flipped and simulated teaching: an exploratory study," *Nurse Education in Practice*, vol. 27, pp. 71–77, 2017.
- [18] W. Xu, W. Xiong, Z. Shao, and Y. Li, "Analysis of effectiveness and performance prediction of sports flipped classroom teaching based on neural networks," *Scientific Programming*, vol. 2021, Article ID 5284457, 7 pages, 2021.
- [19] C. R. Tainter, N. L. Wong, G. A. Cudemus-Deseda, and E. A. Bittner, "The 'flipped classroom' model for teaching in the intensive care unit," *Journal of Intensive Care Medicine*, vol. 32, no. 3, pp. 187–196, 2017.
- [20] L. M. Dooley, S. Frankland, E. Boller, and E. Tudor, "Implementing the flipped classroom in a veterinary pre-clinical science course: student engagement, performance, and satisfaction," *Journal of Veterinary Medical Education*, vol. 45, no. 2, pp. 195–203, 2018.



- [21] F. Yang, W. Lin, and Y. Wang, "Flipped classroom combined with case-based learning is an effective teaching modality in nephrology clerkship," *BMC Medical Education*, vol. 21, no. 1, p. 276, 2021.
- [22] H. Y. Kang and H. R. Kim, "Impact of blended learning on learning outcomes in the public healthcare education course: a review of flipped classroom with team-based learning," *BMC Medical Education*, vol. 21, no. 1, p. 78, 2021.
- [23] J. H. L. Koh, "Four pedagogical dimensions for understanding flipped classroom practices in higher education: a systematic review," *Educational Sciences: Theory and Practice*, vol. 19, no. 4, pp. 14–33, 2019.
- [24] K. Powers, "Bringing simulation to the classroom using an unfolding video patient scenario: a quasi-experimental study to examine student satisfaction, self-confidence, and perceptions of simulation design," *Nurse Education Today*, vol. 86, p. 104324, 2020.
- [25] D. Vlachopoulos and A. Makri, "The effect of games and simulations on higher education: a systematic literature review," *International Journal of Educational Technology in Higher Education*, vol. 14, no. 1, pp. 1–33, 2017.
- [26] A. J. Crothers, J. Bagg, and R. McKerlie, "The flipped classroom for pre-clinical dental skills teaching—a reflective commentary," *British Dental Journal*, vol. 222, no. 9, pp. 709–713, 2017.

## Research Article

# Industry 4.0-Oriented Chipless RFID Backscatter Signal Variable Polarization Amplitude Deep Learning Coding

Guolong Shi<sup>1,2</sup>, Yigang He,<sup>1</sup> Lichuan Gu,<sup>2</sup> and Jun Jiao<sup>2</sup>

<sup>1</sup>School of Electrical Engineering and Automation, Wuhan University, Wuhan, Hubei 430072, China

<sup>2</sup>School of Information and Computer, Anhui Agricultural University, Hefei, Anhui 230036, China

Correspondence should be addressed to Guolong Shi; shiguolong@whu.edu.cn

Received 10 August 2021; Revised 31 August 2021; Accepted 1 September 2021; Published 24 September 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Guolong Shi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Due to the weak network security protection capabilities of control system network protocols under Industry 4.0, the research on industrial control network intrusion detection is still in its infancy. This article discussed and researched the intrusion prevention technology of industrial control networks based on deep learning. According to the electromagnetic scattering theory, the backscatter signal model of the chipless tag was established as a chipless tag structure. Polarized deep learning coding was used for the label; that was, deep learning coding was performed on the copolarization component and the cross-polarization component at the same time, and a 16-bit deep learning coding bit number was obtained. The wave crest deep learning coding was used for the split ellipse ring patch label, and the 6-bit deep learning coding bit number was obtained. Then, the poles of the scattered signal of the tag were extracted to identify the tag. The variable polarization effect was achieved by adopting the dipole resonant unit with the two ends bent. Aiming at the problem of low detection rate caused by the shallow selection of feature classification of intrusion prevention systems, an industrial control network intrusion prevention model based on self-deep learning encoders and extreme learning machines was proposed to extract features from industrial control network data through deep learning. For accurate classification, the theoretical judgment was also verified through simulation experiments, and it was proved that the detection rate of the model has also improved. It forms a set of industrial control network intrusion prevention system with complete functions and superior performance with data acquisition module, system log module, defense response module, central control module, etc. The matrix beam algorithm was used to extract the poles and residues for the late response, and the extracted poles and residues were used to reconstruct the signal. The reconstructed signal was compared with the scattered signal to verify the correctness of the pole extraction. Finally, the tags were processed and tested in the actual environment, and the measured results were consistent with the theoretical analysis and simulation results.

## 1. Introduction

“Industry 4.0” is an era in which information technology promotes industrial transformation. At the same time, the concept of “Internet +” has triggered a major integration and transformation of information technology and the industry [1]. Therefore, the extensive application of IT technology in industrial control systems has transformed it from a proprietary and closed system into a highly open and interconnected system. While work is convenient and production efficiency is improved, the networked development of industrial control systems has led to system safety in case of that risks and threats of intrusion continue to increase [2]. Chip-

less tags do not require silicon chips, so their production costs are relatively low compared to chip tags. Compared with traditional automatic identification technologies such as bar codes, two-dimensional codes, and card recognition, RFID technology has an absolute advantage [3]. As an emerging automatic identification technology, frequency identification has the characteristics of small size, large capacity, long life, and reusability. This technology can be combined with Internet, communication, and other technologies to realize the tracking and information of items on a global scale by M. R. Souryal et al. [4]. With the continuous development of this technology, its application field is expanding day by day, and it has become a current hot

research field with a wide range of application prospects. The cost of labels is a major bottleneck in the application and development of technology. Only by reducing the cost of labels can it be widely used [5]. The design and research of the frequency band chipless or label is based on this background. Because there is no microchip and on-chip power supply, the chipless arsenic label greatly reduces the cost. The smart card has friction with the reader terminal every time it is used, which is easy to wear. The RFID tag and the reader terminal communicate wirelessly, there is no friction, and the RFID tag can be encapsulated. Inside the shell, there is no direct contact with the outside world, and it will not affect the recognition effect, so the RFID tag is more durable [6].

Low-cost chipless radio frequency tags have the most potential to become a substitute for barcodes, so chipless tags have become a hot spot in current research. At present, RFID tags widely used in the commercial market are mainly divided into active tags and passive tags with integrated silicon chips. These two types of tags have a high cost of mass production, which limits the popularization and development of RFID technology [7]. Therefore, the concept of chipless RFID tags with low cost advantages is proposed.

Based on the in-depth understanding of the importance of polarization information, variable polarization technology has begun to be used correspondingly in stealth, antistealth and interference, and antijamming [8]. With the further development of technology, it has put forward higher requirements for the polarization system: small size, low power consumption, and good antioverload performance. Therefore, this paper takes the research of proximity detection system RFID as the background and mainly focuses on the difficulty of miniaturization of variable polarization RFID in variable polarization system [9]. The chipless tag based on phase deep learning coding is composed of three square microstrip patch RFIDs, and each RFID is loaded with an open-circuit microstrip transmission line. The three RFIDs have adjacent resonant frequencies. When excited by their respective resonant frequencies, they reflect backscattered signals with different phase characteristics, which completes the phase modulation of the backscattered signals [10]. Since the symbol period of the tag is controlled by the frequency of the system oscillator, shortening the symbol period requires increasing the oscillator frequency. The higher the oscillator frequency, the greater the energy consumption. It is not wise to increase the data rate by increasing the oscillator frequency. Aiming at the problem of low detection rate caused by the shallow selection of feature classification of intrusion prevention systems, an industrial control network intrusion prevention model based on self-deep learning encoders and extreme learning machines is proposed to extract features from industrial control network data through deep learning. For accurate classification, the theoretical judgment is also verified by simulation experiments, and it is proved that the detection rate of the model has also improved [11]. According to the different phase characteristics of the backscatter signal, each tag can be designed as a unique and immutable chipless tag design. In addition to the deep learning coding capacity and size of the tag, whether the tag can be realized in the actual environ-

ment which is also very important. Since depolarizing technology can improve the robustness of tag detection, the research on variable polarization chipless RFID tags has important application value and practical significance [12].

## 2. Related Work

Low-cost chipless radio frequency tags have the most potential to become a bar code substitute. Experts and scholars have done a lot of work and designed many types of chipless tags. The label design methods are also different. They are classified according to the deep learning coding method and are roughly divided into the following four categories [13]: based on time-domain delayed reflection, the typical label structure is based on surface acoustic wave (SAW) for the spectral characteristics, and its typical tag structure is a chipless tag with a multispiral resonator. Based on the amplitude/phase modulation of the reflected signal, its typical tag structure is a chipless tag. For the conductor complex natural resonance frequency, the principle of this type of label is that the scattered signal of the label carries the structural information of the label, and the purpose of identifying the label is achieved through the analysis of the scattered signal. At present, many types of chipless RFID tags have been designed based on the above four methods. Among them, tags are designed based on frequency domain characteristics, and the complex natural resonance frequency of conductors is more common [14]. So far, many domestic and foreign experts, scholars, and scientific research teams have conducted in-depth research on chipless RFID. These scientific research teams have put forward a variety of chipless RFID tag design ideas, hoping to solve the current chipless RFID tag design from different angles. The size, deep learning coding capacity, and label detection problems are faced by the Internet. Garbati et al. [15] divided chipless RFID tags into the following types according to the different ways of deep learning coding of data information: based on time domain deep learning coding and frequency domain deep learning coding, early polarization studies of type and other types began in the field of radar polarization in the early 1990s. Among them, Bekkali et al. [16] proposed a three-parameter trajectory method. Decarli and Dardari [17] proposed the concept of optimal target polarization, pointing out that there are several polarization states corresponding to the maximum and minimum energy of the radar receiving signal. He proposed the zero-polarization description method, established the basic theory of radar polarization, and became the founder of radar polarization research.

The surface acoustic wave is reflected by the reflector group for deep learning coding, and the reflected surface acoustic wave is emitted through the fork energy converter to generate electromagnetic waves. The RFID tag has a range of up to several meters, can identify high-speed moving targets, and has strong anti-interference performance. However, the fork energy converter is expensive, and the reflector group is large in size and cannot be bent. Aliasgari et al. [18] used two-way cross-energy converters and Z-shaped path reflector groups to greatly reduce the volume of RFID tags. Passive tags based on multiresonators generally consist of

one receiving RFID, one resonant circuit, and one transmitting RFID sister. It needs to transmit a pulse signal with a wide frequency spectrum. In the resonant circuit, each resonant unit is a bit, which receives RFID. The received pulse signal goes through the resonant circuit for deep learning coding and then is transmitted by the RFID transmitter. The label designed by the feeding method can be completely printed, and the performance is stable. Babaeian and Karmakar [19] designed a 35-bit chipless RFID tag, which has a larger capacity, but the circuit is complicated, and the disadvantages of larger size still exist. Guidi et al. [20] designed a 3-bit chipless RFID tag with 3 patch units with close resonant frequencies and a rate close to the open-circuit high-impedance line. Changing the length of the high-resistance line is equivalent to changing the load of the chip RFID. Thus, affecting the phase of the backscattered signal, so the deep learning coding capacity is small, and the size is too large. Some scholars have designed a circular patch label with a circular slit. The radius of the circular slit has a one-to-one correspondence with the wavelength of the resonant frequency of the backscatter signal, so this kind of label has many advantages such as large deep learning coding capacity and small size, but the structure is relatively complex [21]. The horizontal slot length of the label patch corresponds to the civil resonance frequency of horizontally polarized electromagnetic waves, and the vertical slot length of the label patch corresponds to the RFID resonance frequency of vertically polarized electromagnetic waves.

This article introduces the basic theory of polarization state representation and elliptical polarization synthesis and discusses and summarizes some methods of polarization parameter measurement. Based on these theoretical knowledge, the principles of receiving and transmitting in the variable polarization system are further discussed, and the basic implementation scheme of the miniaturized variable polarization system is given. Finally, based on the miniaturization requirements of the variable polarization system, an important part of the system, a double-end-fed square microstrip dual-polarization RFID, is simulated and designed. Based on the principle of backscatter modulation, a method based on the polarization characteristics of electromagnetic waves is proposed. Chipless RFID system with variable polarization technology and the indicators of the system and each component are given. The system has the advantages of large deep learning coding capacity, simple structure, and low cost. And its variable polarization characteristics are simulated, and the results ideally verify that the active variable polarization system can use dual-end-fed dual-polarized microstrip RFID as an orthogonal dual-fed RFID to achieve the composite output of various polarized waves.

### 3. Design of a Chipless RFID System with Variable Polarization of Backscattered Signal and Variable Polarization for Industry 4.0-Oriented Amplitude Deep Learning Coding

*3.1. RFID Performance Mechanism.* RFID is the main component of the radio frequency system. For the identification

system, RFID should have a simple structure and low cost. The main radio frequency RFIDs currently used include monopole RFID, dipole RFID, and microstrip RFID. Figure 1 shows the hierarchical distribution of RFID performance mechanisms.

For RFID in the wireless communication system, the working principle follows the electromagnetic field theory; through the alternating electromagnetic field to form an electromagnetic wave to transmit information, the electromagnetic field propagation under ideal conditions satisfies Maxwell's equation.

$$\nabla \times E + j\omega\alpha \times H = 0, \quad (1)$$

$$\nabla \times H + j\omega\alpha \times E = J. \quad (2)$$

It can be seen from the above formula that the propagation direction of the electromagnetic field is consistent with the direction of the energy flow density. According to the electromagnetic field theory, the electromagnetic field area which was emitted by RFID in the radio frequency identification system can be divided into three parts: induction near field, radiation near field, and far field. The induction near field is the area close to the radio RFID and dominated by the induction field. The main form of energy storage is stored; the box radio near field is mainly dominated by the radio field, and the radiation angle is related to the distance of the RFID. The energy in the far field of the box is mainly emitted by electromagnetic waves, and the radiation angle has nothing to do with the distance. The RFID in this article belongs to ultrahigh frequency RFID, which mainly works in the far field of radiation.

$$E = idl/4\pi \times \cos \alpha \times \exp(-jw). \quad (3)$$

RFID is a device for sending and receiving signals in a radio frequency identification system. The performance of RFID will affect the identification of tags. This section mainly introduces several main performance parameters of RFID. The relationship between RFID's box shootability and space coordinates is called the RFID directional function, and the national shape drawn according to the functional relationship is called the directional diagram. For the radiation characteristics of RFID, the directivity function is defined as:

$$\iint L/4\pi r^2 \times \sin \alpha \cos \beta \times d\alpha d\beta = H. \quad (4)$$

Among them,  $r$  is the radiation power,  $L$  is the input power, and  $H$  is the total power consumed in the RFID conductor and the dielectric plate. If the power consumed by RFID is regarded as the loss of resistance, so:

$$\begin{cases} S_x = Axa^2/4\pi \times P(\text{tag}), \\ S_y = Ayb^2/4\pi \times G(\text{tag}). \end{cases} \quad (5)$$

Therefore, the relationship between the efficiency of RFID and the radiation resistance and loss resistance can

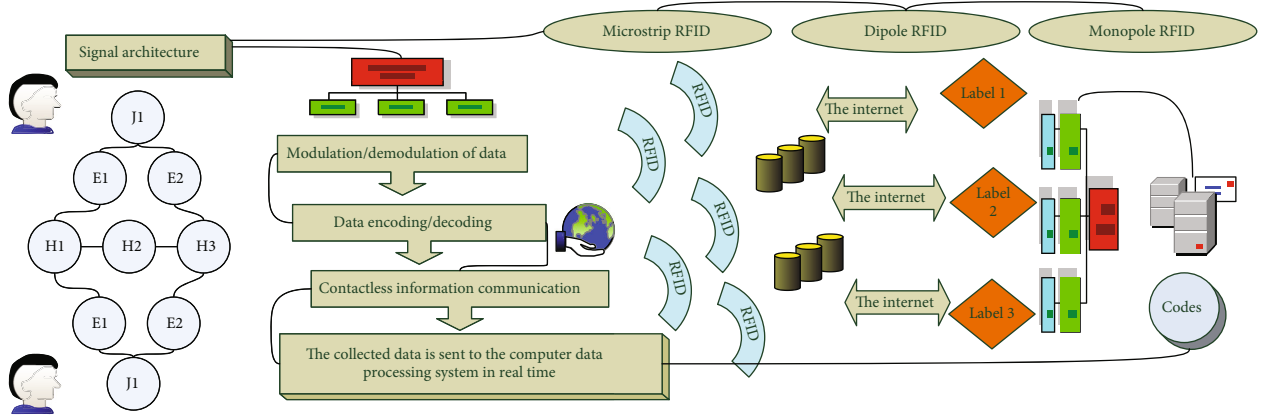


FIGURE 1: Hierarchical distribution of RFID performance mechanism.

be obtained. Therefore, the RFID efficiency can be improved by reducing the loss resistance and increasing the box emission resistance. The input impedance of RFID is related to the shape, size, and excitation method of RFID and the distance between surrounding objects. RFID is connected to the feeder, and impedance matching is usually required. When the input impedance of RFID is pure resistance and the characteristic impedance of the feeder is equal, no power reflection is the best state at this time.

$$I = \begin{bmatrix} i \cos \alpha \sin \beta & 0 & 0 \\ 0 & i \sin \alpha \cos \beta & 0 \\ 0 & 0 & i \cos \alpha \cos \beta \end{bmatrix}. \quad (6)$$

RFID polarization represented the trajectory of the end point of the space electric field vector. The polarization mode of RFID can be divided into  $H$  kinds of linear polarization, circular polarization, and column circular polarization. Since the length of a monopole is about a quarter of a wavelength and its working frequency is determined by the size of its monopole, RFID of various working frequency bands can be made as needed.

$$\frac{w_r + 1}{2} + \frac{w_r - 1}{2} \times \frac{w_r^2}{a} = W, \quad (7)$$

$$\frac{E'}{H'} = \left| \frac{W_1 - W_a^*}{W_1 + W_a^*} \right|^2 \times \exp(-jw). \quad (8)$$

For the shape of the RFID box radiation pattern, it is omnidirectional on the  $H$  side; it is a figure eight on the  $E$  side. For the RFID feeding method, the input impedance of this type of RFID is close to 50 ohms, which matches the 50 ohms feeder line. Therefore, microstrip lines, coaxial cables, coplanar waveguides, etc., can be used for feeding. The various performances of single-pole RFID make it applicable to RFID systems.

**3.2. Variable Polarization Backscatter Modulation.** The backscatter modulation terminal is responsible for processing the received signal and reading the characteristic infor-

mation of the object. It is composed of middleware and a background processing network. The middleware is responsible for communicating with the reader, obtaining the information from the reader and performing corresponding processing, and at the same time transmitting the information to the back-end network for processing, and finally obtaining the characteristic information of the object. The background processing network generally includes modules of data center, control center, and network equipment. The data must be used to store the characteristic data of the item. After receiving the information from the reader, the background network will query the database according to the item ID to determine the characteristic data of the item. Generally speaking, the close staring system is mostly realized by quasi-static field collision. Figure 2 shows the variable polarization backscatter modulation mechanism. As mentioned earlier, the energy exchange mode between the reader and the tag at this time is similar to the transformer model, that is, the load modulation mode. This method actually changes the voltage on the reader's antenna by changing the on and off of the load resistance on the tag antenna, so as to realize the amplitude modulation of the antenna voltage by the short-distance tag. This modulation method is widely used in the control of system and mainly controls the background processing network to ensure normal communication.

For the staring system of 2 GHz or higher frequency, it works in the typical far field. The energy transfer method between the reader and the tag can only be backscatter modulation. Backscatter modulation refers to the communication method used when the electronic tag in the passive staring system sends data to the reader. The tag generally contains an antenna and a chip. An impedance switch is used in the chip to adjust the input impedance of the chip itself. Since the data type of the industrial control network intrusion detection data set is a mixture of continuous and discrete data, it is necessary to perform preprocessing such as normalization of the data. The sparse autoencoder can make the preprocessed data more sparse after feature extraction. The feature data with high sparsity after dimensionality reduction is used as the input of the classifier, and then, the classifier is used to classify the feature data. Assuming that

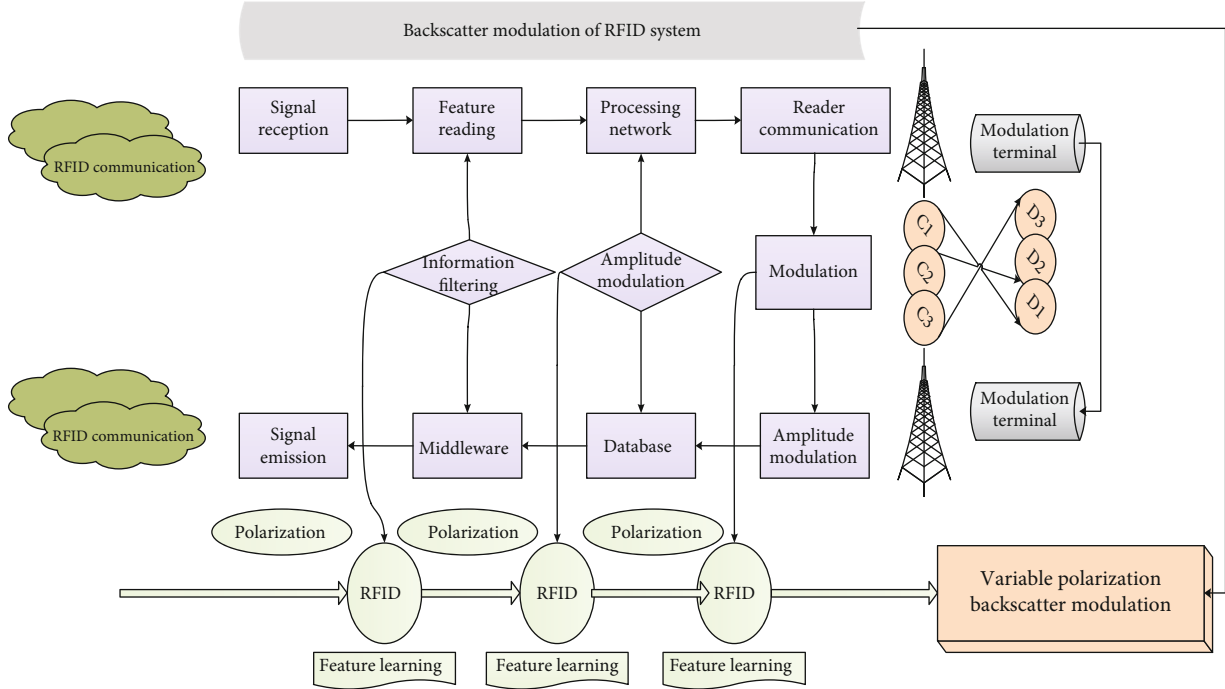


FIGURE 2: Variable polarization backscatter modulation mechanism.

the data signal to be sent is a signal with two levels, it is modulated by a logic gate and an intermediate frequency signal, and the modulated signal is controlled. The opening and closing of the impedance switch change the input impedance of the chip, that is, changes the impedance matching between the chip and the antenna. When the tag antenna is conjugated to the connected tag chip, most of the energy received by the antenna will be transferred to the chip, and the energy reflected back to the reader is small. The reflected echo is modulated in a method similar to RFID modulation.

**3.3. Amplitude Deep Learning Coding Optimization.** Since the deep learning coding angle of the linear slot label is related to the direction of the excitation electric field, the strip gap in the linear slot label structure is changed to a shape, showing a relationship where the deep learning coding angle is not related to the direction of the excitation electric field. Figure 3 shows a schematic diagram of amplitude deep learning coding optimization. The V-shaped structure contains two slit arms A and B of equal size, and the size of the two arms is electrode. The fixed arm A is placed in the horizontal direction, and the angle between the two arms is 0 for deep learning to encode data. The horizontally polarized plane electromagnetic wave with electric field intensity enters the new label vertically. It can be seen that the current flows are two circuits around the gap. The formed gap label is used as a standard label, and the scattered electric field component of arm A in the horizontal direction is E. The value of the electric field component in the horizontal direction and the component value in the vertical direction of the scattered electric field around the arm B are obtained, respectively.

According to the distribution of the frequency spectrum when the label is working, we select 16 angles with an interval of 10 for the deep learning coding address information, corresponding to the 4-bit address information, and the deep learning coding capacity is 4 bit. Substituting the electric field value of the label at the working frequency into the formula, the recognition result is shown in paper.

It can be seen that all addresses can be distinguished correctly, and the angle recognition error is less than 2.24. And different deep learning coding states all occupy the same frequency band range of 2.92 GHz-4.28 GHz. Figure 4 shows the broken line graph of the amplitude information of the deep learning encoding. In passive systems with frequency bands or higher frequencies, the electromagnetic waves emitted by the reader are reflected to complete the data transmission from the tag to the reader.

The radio frequency signal generated by the reader is radiated to the space by the antenna. Assuming that the gain of the reader's transmitting antenna is taste, the transmitting power is the effective radiation power of the antenna, which is the product of the reader's transmitting power and the antenna gain. The pooling layer can effectively reduce the size of the matrix by reducing the length, width, and depth of the matrix, thereby reducing the parameters in the subsequent fully connected layer and at the same time making feature sparse can accelerate the calculation speed and prevent overfitting. The forward propagation of the pooling layer is completed by performing a maximum or average operation on the movement of the pooling core. According to radar technology, an object whose volume exceeds half of the electromagnetic wavelength can be reflected, and the energy reflected by the tag is proportional to the reflective cross-sectional area of the object. In addition, considering the

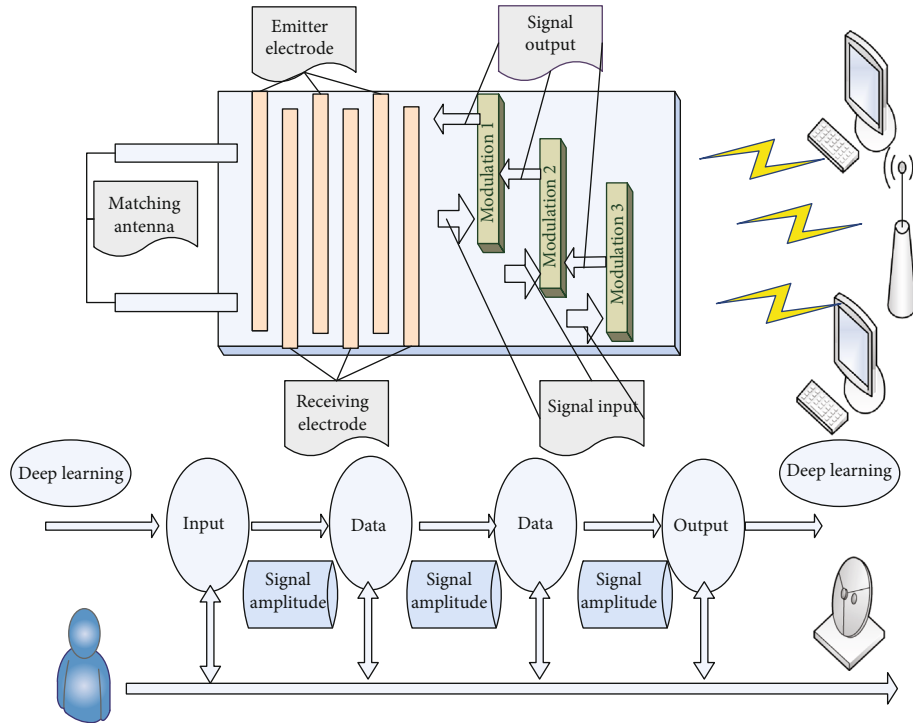


FIGURE 3: Schematic diagram of amplitude deep learning coding optimization.

impedance matching between the tag chip and the tag antenna, the impedance matching coefficient of the tag is defined as  $q$ . Considering the limit situation, when the tag chip and the antenna impedance match, all the energy is transmitted from the tag antenna to the inside of the chip, and the tag reflects the power. When the tag chip and the antenna impedance do not match, all the energy is reflected back to the reader, and the system uses this backscatter modulation method to return the signal to the reader.

#### 4. Industry 4.0-Oriented Amplitude Deep Learning Coding Backscatter Signal Variable Polarization Chipless RFID System Application and Analysis

**4.1. Backscatter Signal Preprocessing.** In this paper, the resonant frequencies of the three antenna units are designed as  $f_1$ ,  $f_2$ , and  $f_3$ , which are all between 2.1 and 2.5 GHz, and the frequency separation between each other is about five stones. The reader in this system has already learned these three frequencies. When the system is working, the reader transmits the frequency of five multifrequency access signals. Here is a simplified model, and the multifrequency access signal is set as a sine wave. The three antenna elements in the chipless tag are excited by their respective resonant frequency waves and reflect backscattered signals with different phase characteristics. The signals with phase characteristics returned by simulation can be obtained. After fixing the phase, we can get three discrete phase values. If each patch unit can be designed to make  $n$  different situations of its own resonant frequency wave in the reflected signal, then

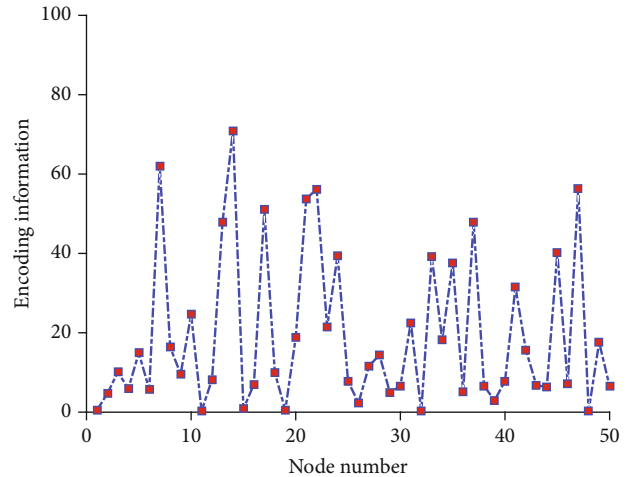


FIGURE 4: The broken line graph of the amplitude information of the deep learning coding.

the chipless tags composed of three patch units have different combinations of minerals. Figure 5 shows the flow chart of the backscatter signal preprocessing. According to the different phase characteristics of the backscattered signal, each tag can be designed to be unique and immutable.

Here, we choose a material with a dielectric constant of  $a$  and a thickness of  $b$ . The working frequency band is between one. According to the design method of the rectangular microstrip antenna introduced above, the side length of the square microstrip patch is initially selected. Without loading any transmission lines, we design the antenna to match its impedance. In this section, commercial electromagnetic simulation software will be used to obtain the parameter values

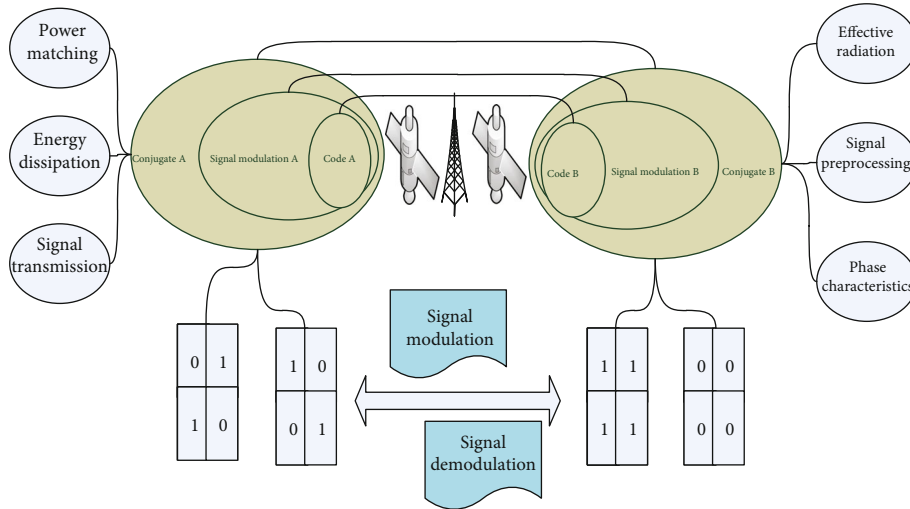


FIGURE 5: Flow chart of backscatter signal preprocessing.

of the optimized square microstrip patch antenna. The characteristic impedance of the embedded microstrip feeder and the width of the feeder are one. The size of the parameters  $a$  and  $b$  can be adjusted to optimize the antenna to match. In the antenna reflection coefficient with the open end of the embedded microstrip feeder as the antenna input port, it can be seen that the antenna is well matched to 50 ohms. Here, we use three square microstrip units to form an antenna array. The three square microstrip patch antenna units have different resonant frequencies, and the resonant frequencies are adjacent, all within a range. We select the appropriate side length according to the design theory of the microstrip patch antenna and the operating frequency requirements.

**4.2. Simulation of Chipless RFID System.** The measurement process is carried out in the actual environment rather than in the microwave anechoic chamber. The measuring instrument is a N5230A vector network analyzer with a transmitting power of 3dB. Two identical standard butterfly antennas are used as the transmitting and receiving antennas of the reader. We connected the two identical standard antennas before measurement. On port 1 and port 2 of the network analyzer, they were used as transmitting and receiving antennas, respectively. The antenna works in the frequency range of 1.572-3.714GHz. We placed the two antennas horizontally and vertically and placed a device that can fix the sensor at a distance of 0.8m from the antenna and placed the device in the middle of the two antennas. It could be seen that the sensor was at 50. Through random selection, 5000 sets of data in the data set are used as the training set, and 1000 sets of data are used as the test set. The sigmoid function is used as the activation function of the sparse autoencoder, the sparsity parameter is 0.25, the number of hidden layers is  $n = 14$ , and the number of cycles is 50. The extreme learning machine as the classification module also uses the sigmoid activation function. After many experiments, the input weight, the number of nodes, the bias of the hidden unit, and other parameters are deter-

mined. The working frequency was the same, and the working frequency band was relatively narrow, which could effectively improve the high spectrum utilization rate. We put the values of the resonance frequency into the formula and saw the article for the recognition results of angle and underresistance. It could be seen that the recognition errors of angles were all less than 3, and the recognition errors of resistors were all less than 10, so the sensor could realize the correct identification of angle 0 and resistance value.

The sensor had two resonant frequencies, and the resonant frequency was only related to the length of the scattering unit and did not change with the angle. Figure 6 shows the amplitude response curve of the resonant frequency of the chipless RFID. It could be seen from the figure that at the resonance frequencies of 15 Hz and 95 Hz, the scattered electric fields in the horizontal and vertical directions were pulled, respectively. Bringing the parameters into the formula, the recognition angles were 91.5 and 101.5, respectively. It could be obtained that the angle errors are 1.5, 1.5, 0.9, and 4.

**4.3. Example Application and Analysis.** The plane wave is incident on the sensor vertically, and the two resonance frequencies are obtained by simulation. At the resonance frequency, the horizontal and vertical backscattered electric fields are  $E_1$ ,  $E_2$ ,  $H_1$ , and  $H_2$ . Since the angle recognition formulas of two scattering units and one scattering unit are the same,  $W$  incorporates the simulated backscattering electric field value into the formula, and  $W$  can realize the recognition of angle from the perspective of deep learning to encode address information. Figure 7 shows the identification curve of the tag angle information under the RFID backscattering electric field. The blue data line is the control group, which corresponds to the algorithm proposed in Reference [20], and the red data line is the experimental group, which corresponds to the algorithm in this paper. It can be seen that the red data line is smoother, which means that the algorithm in this paper is more superior. The respective resonant frequencies of the antenna are  $f_1$ ,  $f_2$ , and  $f_3$ , and



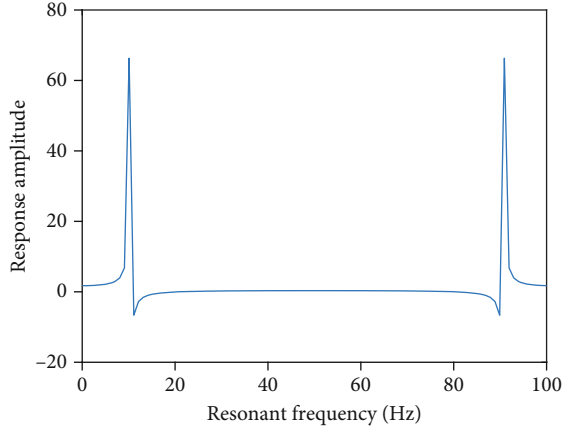


FIGURE 6: Amplitude response curve of resonant frequency of chipless RFID.

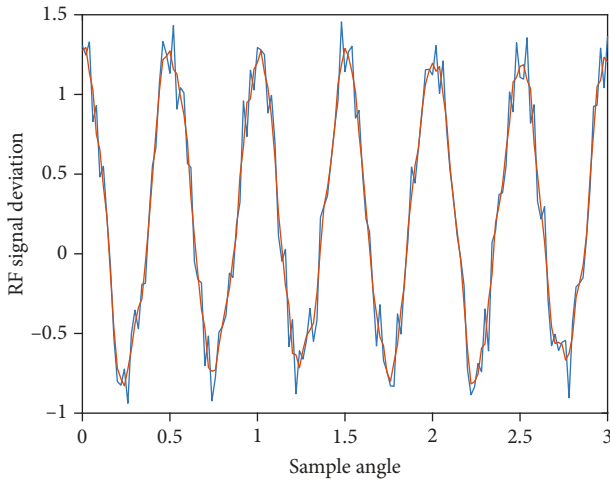


FIGURE 7: The identification curve of tag angle information under RFID backscattering electric field.

the return loss deviation value of each resonant point is lower than 1.5, which can be well matched. And it can be seen from the paper that the maximum mutual coupling between the three antennas is as low as -20 dB, which shows that the mutual coupling between the three antenna elements is small, and the mutual influence is within an acceptable range.

By measuring the horizontal and vertical components of the backscattered electric field, the identification of the tag's angle information is realized. The tag deep learning coding capacity can be up to 4 bit, the angle recognition error of the linear slot label is less than 1.55, and the angle recognition error of the V-shaped slot tag is less than 2.24. The measurement result of the tag is consistent with the simulation result, and the measured resonance point is slightly smaller than 5 GHz. The designed RFID model has a simple structure and low cost. Different deep learning coding tags occupy the same frequency spectrum and occupy a narrow frequency band, which can reduce the cost of the entire radio frequency identification system. The designed tag can

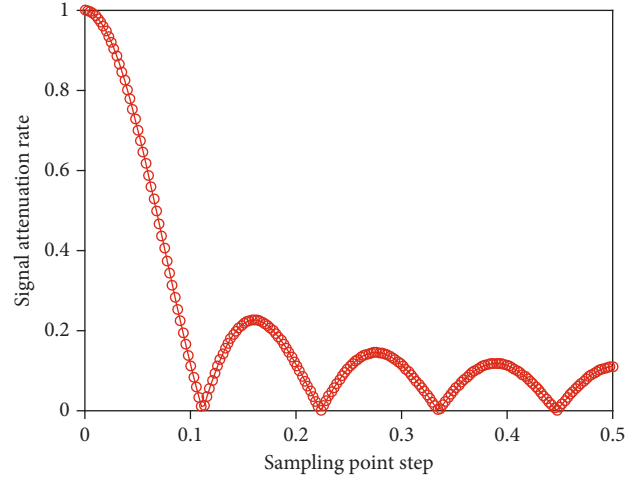


FIGURE 8: RFID polarization excitation signal gain rate depends on the sampling point of resonance frequency.

be applied to the Internet of Things as an angle sensor. When RFID frequency is low, the antenna feed mode is traditional coplanar waveguide feed; the antenna has two resonance points of 3.6 GHz. As it increases, the direction changes little, and it moves to the left. It is approximately equal to a quarter of the guided wave wavelength corresponding to 2. Figure 8 shows the dependence of the RFID polarization excitation signal gain rate on the sampling point of the resonance frequency. For the entire antenna, the asymmetry of ACPW (asymmetric coplanar waveguide) causes the antenna to have four resonance points. By adjusting signal mode can be superimposed together to obtain a broadband antenna. However, due to the appearance of another butterfly-shaped gap, RFID attenuation rate will shift a little, and the signals merge into a resonant frequency.

The sensor is excited by linearly polarized waves, and the placement direction of the fixed arm  $A$  is the same as the polarization direction of the excitation electromagnetic wave. That is, when the excitation electromagnetic wave is a horizontally polarized wave, the fixed arm  $A$  is placed in the horizontal direction. The results show that after the training set has been continuously increased the number of iterations, the improved sparse autoencoder reduces the loss value of the training set after dimensionality reduction and stabilizes at 0.1. In other words, the training network is continuously converging, and the algorithm is extracted for the effect of the feature also tends to stabilize. When the wave is vertically polarized, the fixed arm  $A$  is placed in the vertical direction. The antenna impedance and magnitude are different, which can be seen that the influence of  $A$  is less than  $1/3$ . With the increase of phase, the degree of change in the imaginary part of impedance in the range of 0.01-0.1 rad/s is greater than that at 100-1000 rad/s. When frequency is less than 0.1 rad/s, the imaginary part of the impedance is lower than 50 dB, and the real part is close to 50.

Figure 9 shows the phase and amplitude information of the scattered electric field at different variable polarization frequencies. Using wireless radio frequency technology to read the sensor's scattered electric field information, we

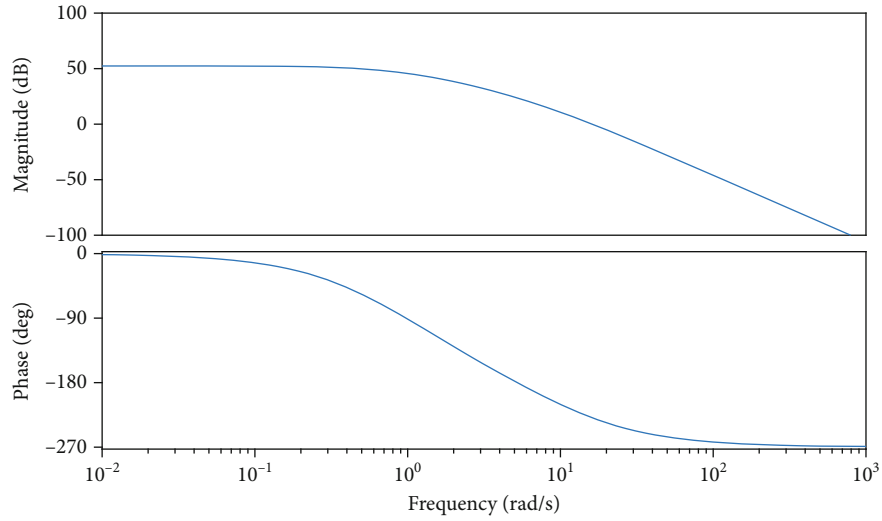


FIGURE 9: Scattered electric field phase amplitude information at different variable polarization frequencies.

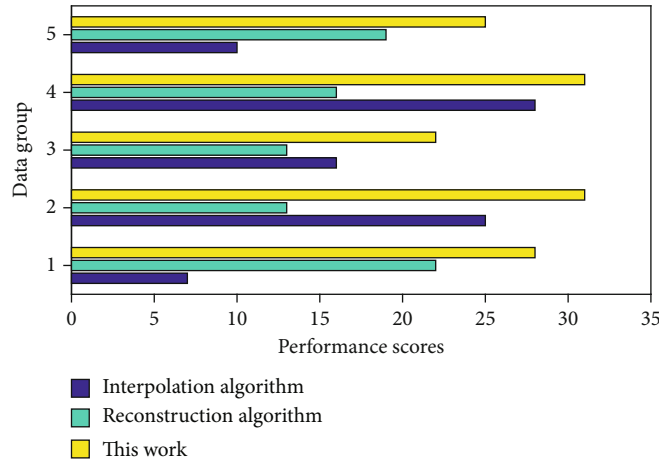


FIGURE 10: Histogram of score distribution of different coding algorithm models.

can realize the identification of the sensor’s address information and the resistance to be measured. Through simulation, it can be known that the resonance frequency of the sensor is only related to the length of the box-fired patch, and the angle is only related to the backscattered electric field of the corresponding radiation unit. By measuring the horizontal and vertical backscattered electric fields, the angle and identification of resistance are given.

Figure 10 shows the histogram of the score distribution of different coding algorithm models. It can be seen from the histogram that the algorithm proposed in this paper has obtained relatively high scores in the 5 experimental groups and has excellent performance compared to the other two algorithms. The yellow bar graph represents the experimental group, and the other two colors of data represent the control group. Compared with interpolation algorithm and reconstruction algorithm, the algorithm designed in this paper performs better. For the resistance sensor with a single type unit, the angle recognition error is less than 3, the resistance value recognition error is less than 1.5, and the deep

learning coding capacity can reach 4 bit. For the resistance sensors of two units, the angle recognition error is less than 2.5, and the resistance recognition is less than 2.0. An RFID mode unit is added to realize the 5-bit deep learning coding capacity. In addition, the RFID sensor in this paper can be directly applied to the radio frequency identification system by changing the address information and resistance value to a wireless signal.

### 5. Conclusion

The components designed in this paper construct a chipless RFID field system based on the variable polarization scheme, and we use  $2 \times 2$  array tag RFID and  $4 \times 4$  array tag RFID to test and analyze the horizontal polarization branch and vertical polarization branch of the system separately. We compare the test results and use  $4 \times 4$  array tag RFID to test and analyze the horizontal polarization branch and vertical polarization branch of the system at the same time. Combined with the measured experimental data, the specific

process of the system identification tag RFID polarization mode is given, and the feasibility of the chipless RFID system based on variable polarization technology is verified. By lengthening and bending one of the feeding slots, an asymmetric coplanar waveguide feeding mode is formed, which improves the impedance matching of the antenna and increases the bandwidth of the antenna. At the same time, the  $H$ -angle loop conduction band is loaded in the butterfly slot, which improves the low-frequency gain of the antenna and makes the gain of the antenna in the working frequency band more flat. The gain within the bandwidth is all higher than 1.5 dB, and the highest gain is 5.53 dB; the simulation and actual measurement results are basically in agreement. The two resonance peaks in the cross-polarization direction are both used to determine the humidity change, which is beneficial to reduce the humidity detection error in practical applications. The two anomaly detection algorithms based on deep learning proposed in this paper are used as signal detection modules, which are combined with data acquisition module, system log module, defense response module, and central control module. Based on the above theoretical knowledge, an intrusion detection algorithm based on autoencoder and extreme learning is established. Through algorithm description and simulation experiments, it is proved that it meets the reliability requirements of "high detection rate and low false alarm rate" of industrial control network. A set of industrial control RFID network signal prevention system with complete functions and superior performance has been formed, and its effectiveness and reliability have been proved through system testing. In addition, the tag has variable polarization characteristics, which can improve the robustness of the sensor's detection in the actual environment.

### Data Availability

The data used to support the findings of this study are included within the article.

### Conflicts of Interest

The authors declare that they have no conflicts of interest.

### Acknowledgments

This work was supported by the China Postdoctoral Fund General Project (No. 2021M692473) and the Natural Science Foundation of Anhui Province (2108085QF260), the National Natural Science Foundation of China Project: "Research on the Integrated Control Method of the Attachment Limit State of Agricultural Tracked Robot Based on Non-stationary Constraint" (No. 31671589), and the Provincial Science and Technology Major Special Project: "Research on Key Technologies for Precise Control of Healthy Pig Breeding and Intelligent Disease Early Warning in a Big Data Environment" (No. 201903a06020009).

### References

- [1] C. Feng, W. Zhang, L. Li, L. Han, X. Chen, and R. Ma, "Angle-based chipless RFID tag with high capacity and insensitivity to polarization," *IEEE Transactions on Antennas and Propagation*, vol. 63, no. 4, pp. 1789–1797, 2015.
- [2] C. Boyer and S. Roy, "Coded QAM backscatter modulation for RFID," *IEEE Transactions on Communications*, vol. 60, no. 7, pp. 1925–1934, 2019.
- [3] J. D. Griffin and G. D. Durgin, "Complete link budgets for backscatter-radio and RFID systems," *IEEE Antennas and Propagation Magazine*, vol. 51, no. 2, pp. 11–25, 2020.
- [4] M. R. Souryal, D. R. Novotny, D. G. Kuester, J. R. Guerrieri, and K. A. Remley, "Impact of RF interference between a passive RFID system and a frequency hopping communications system in the 900 MHz ISM band," *IEEE Electromagnetic Compatibility Magazine*, vol. 1, no. 3, pp. 97–102, 2012.
- [5] U. Mc Carthy, G. Ayalew, F. Butler, K. McDonnell, and S. Ward, "The effects of item composition, tag inlay design, reader antenna polarization, power and transponder orientation on the dynamic coupling efficiency of backscatter ultra-high frequency radio frequency identification," *Packaging Technology and Science: An International Journal*, vol. 22, no. 4, pp. 241–248, 2009.
- [6] P. V. Nikitin, K. V. S. Rao, and LabVIEW-based UHF, "RFID tag test and measurement system," *IEEE Transactions on Industrial Electronics*, vol. 56, no. 7, pp. 2374–2381, 2020.
- [7] J. Zhang, G. Y. Tian, A. M. J. Marindra, A. Sunny, and A. Zhao, "A review of passive RFID tag antenna-based sensors and systems for structural health monitoring applications," *Sensors*, vol. 17, no. 2, p. 265, 2017.
- [8] C. Herrojo, F. Paredes, J. Mata-Contreras, and F. Martín, "Chipless-RFID: a review and recent developments," *Sensors*, vol. 19, no. 15, p. 3385, 2019.
- [9] E. Denicke, H. Hartmann, N. Peitzmeier, and B. Geck, "Backscatter beamforming: a transponder for novel MIMO RFID transmission schemes," *IEEE Journal of Radio Frequency Identification*, vol. 2, no. 2, pp. 80–85, 2018.
- [10] V. R. Sajitha, C. M. Nijas, T. K. Roshna, R. Vivek, K. Vasudevan, and P. Mohanan, "Polarization independent chipless RFID tag," *Microwave and Optical Technology Letters*, vol. 57, no. 8, pp. 1889–1894, 2015.
- [11] C. Herrojo, J. Mata-Contreras, F. Paredes, A. Núñez, E. Ramon, and F. Martín, "Near-field chipless-RFID tags with sequential bit reading implemented in plastic substrates," *Journal of Magnetism and Magnetic Materials*, vol. 459, pp. 322–327, 2018.
- [12] M. A. Islam and N. C. Karmakar, "Real-world implementation challenges of a novel dual-polarized compact printable chipless RFID tag," *IEEE Transactions on Microwave Theory and Techniques*, vol. 63, no. 12, pp. 4581–4591, 2020.
- [13] M. Zomorodi, N. C. Karmakar, and S. G. Bansal, "Introduction of electromagnetic image-based chipless RFID system," *Sensor Networks and Information Processing*, vol. 3, pp. 443–448, 2019.
- [14] V. Mulloni and M. Donelli, "Chipless RFID sensors for the Internet of Things: challenges and opportunities," *Sensors*, vol. 20, no. 7, p. 2135, 2020.
- [15] M. Garbati, E. Perret, R. Siragusa, and C. Halope, "Ultrawideband chipless RFID: reader technology from SFCW to IR-UWB," *IEEE Microwave Magazine*, vol. 20, no. 6, pp. 74–88, 2019.

- [16] A. Bekkali, S. Zou, A. Kadri, M. Crisp, and R. Penty, "Performance analysis of passive UHF RFID systems under cascaded fading channels and interference effects," *IEEE Transactions on Wireless Communications*, vol. 14, no. 3, pp. 1421–1433, 2015.
- [17] N. Decarli and D. Dardari, "Time domain measurements of signals backscattered by wideband RFID tags," *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 11, pp. 2548–2560, 2018.
- [18] J. Aliasgari, M. Forouzandeh, and N. Karmakar, "Chipless RFID readers for frequency-coded tags: time-domain or frequency-domain?," *IEEE Journal of Radio Frequency Identification*, vol. 4, no. 2, pp. 146–158, 2020.
- [19] F. Babaeian and N. C. Karmakar, "Hybrid chipless RFID tags—a pathway to EPC global standard," *IEEE Access*, vol. 6, pp. 67415–67426, 2018.
- [20] F. Guidi, N. Decarli, D. Dardari, C. Roblin, and A. Sibille, "Performance of UWB backscatter modulation in multi-tag RFID scenario using experimental data," *IEEE Ultra-Wideband*, vol. 1, pp. 484–488, 2011.
- [21] M. Forouzandeh and N. Karmakar, "Self-interference cancellation in frequency-domain chipless RFID readers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 67, no. 5, pp. 1994–2009, 2019.

## Research Article

# 5G Joint Artificial Intelligence Technology in the Innovation and Reform of University English Education

**Xia Sun** 

*Department of Foreign Language, Hefei Normal University, Hefei, Anhui 230061, China*

Correspondence should be addressed to Xia Sun; [lookingforsummer@126.com](mailto:lookingforsummer@126.com)

Received 15 August 2021; Revised 2 September 2021; Accepted 3 September 2021; Published 23 September 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Xia Sun. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper considers the issue of human subjectivity in the system of “5G + AI + Education” from the perspective of, on the one hand, the real need for the problems that gradually emerge in the new round of development and application of artificial intelligence, and a philosophical reflection on the application of artificial intelligence in specific fields, on the other hand. It is also a further examination of the issue of human subjectivity in the new context. On the other hand, it is also a further examination of the issue of human subjectivity in the new context of the times, which can also provide students with an immersive learning environment, and AI artificial intelligence and hologram technology can enhance students’ motivation. This paper shows the specific steps and implementation measures of “5G” technology into online oral teaching and provides a case study design to explore the new online oral teaching model, summarizing the advantages and proposing solutions to the shortcomings. The system visualizes each step of gesture recognition to facilitate students’ understanding. Students can experience the process of gesture recognition according to the guidance of the interactive interface, and then, the complex and abstract gesture recognition process is explained with a figurative example, which is conducive to primary and secondary school students’ deeper understanding and improved logical thinking. This will help primary and secondary school students to have a deeper understanding and improve their logical thinking skills. Finally, a comparison experiment is designed to verify the effectiveness of using this system to learn AI knowledge compared with traditional learning methods. The experimental results are analyzed to prove that using this system to learn AI knowledge is effective and helps improve users’ interest in learning and hands-on ability.

## 1. Introduction

In recent years, with the rapid spread of AI technology in life, production, and learning, our society is currently accelerating into such an era, where people and AI work and live together in an AI era. In the context of such an era, it is a matter of urgency to develop AI education to provide students with some exposure and understanding of AI and to better adapt to working and learning in today’s smart life. Most international students from around the world were unable to return to school to conduct classes, which meant that online instruction became the primary, if not the only, method of instruction for teaching majors [1]. However, the online speaking courses during the epidemic also revealed many problems, such as the different situations of the teaching audience, the complexity and variety of teach-

ing platforms, and the difficulty in conducting classroom exercises. Online teaching is difficult for students to feel the real communication environment, and some students may be afraid to speak in class due to lack of confidence, or even never participate in practice. Teachers are unable to get a real sense of the student’s state of mind, communication is difficult, and the quality of teaching is greatly compromised. Teachers are faced with unprecedented challenges when teaching spoken Chinese online [2]. As time progresses, teaching methods need to be constantly updated. The goal of the language is communication, and the traditional mode of teaching oral language can no longer meet the teaching needs of teachers and the learning needs of learners. By using 5G technology to develop a new model of teaching oral language, we can improve learners’ motivation and help them master the skills of speaking Chinese

better on the one hand, and teachers can update teaching materials and enrich classroom content to achieve the teaching purpose more easily and quickly on the other hand.

In recent years, with the continuous improvement of deep learning, algorithm models gradually improved, the circulation and sharing of data resources, not only to make artificial intelligence in all occupations, the process of landing applications significantly accelerated, but also formed a strong thrust to promote the transformation and upgrading of traditional industries. For the transformation and upgrading of traditional industries to inject a strong thrust [3] are by their accumulation of massive data so that artificial intelligence in the commercial application of the big show. For schools, the application of artificial intelligence will change the existing school-running form, expand student learning space, help form a student-centered learning environment, and help alleviate the shortage of excellent teachers and the lack of English teaching resources in impoverished areas. For teachers, artificial intelligence technology has the advantages of personalized and accurate testing of English proficiency, real-time and accurate correction of English pronunciation, flexible and intelligent assessment of oral ability, and efficient and accurate recording of teaching data [4]. It can assist teachers in preparing lessons and remove teachers from simple and repetitive tasks such as correcting homework. It is freed from the work of the students to implement personalized guidance and evaluation for students, thereby improving the quality of teaching and improving teaching efficiency. For students, the emergence of artificial intelligence products for English learning has changed the traditional learning methods that students can only obtain knowledge from teachers and textbooks; on the other hand, it provides students with more personalized learning channels and learning methods, so personalized learning has gradually become the mainstream form, thereby improving students' learning experience and helping students learn English faster and better [5].

The main purpose of this research paper is to address the needs of social practice. It is well known that many mature disciplines have little potential for valuable research. Educational technology is precisely an immature discipline. Educational technology itself is multidimensional, developed based on a synthesis of multidisciplinary theories and technologies, and it has great potential. However, existing research has been zigzagging forward, and some even say that research related to the discipline has stopped, which is the process of zigzag development that the discipline is bound to go through. The in-depth reflection on the prospects of educational technology development reflects a profound exploration of the discipline. Educational technology is a discipline that studies the nature, concept, formation, and development, and types of educational technology as its research object. Educational technology has existed since the first-day education was created. However, only when educational technology develops to a certain stage does the discipline needed by society—educational technology—gradually emerge. The development of the discipline of educational technology is closely related to the development of technology, and the research content of the discipline of educational

technology will change accordingly with technological progress and social development, and it will also be influenced by the changing needs of human beings. Educational technology has a long history of development, is an important aspect of educational activities, and is playing an increasingly important role in an ever-changing society. This is based on the results of the continuous development of modern science and technology, which is what has given educational technology a richer connotation.

## 2. Current Status of Research

Educational artificial intelligence is a new research field that investigates the integration of artificial intelligence with educational activities, and the goal it pursues is to create conditions for student learning by observing and understanding the learning process [6]. Austin pointed out that educational artificial intelligence focuses on better student engagement through computers and teaching platforms to help teachers teach more effectively [7]. In addition, some scholars have discussed the application of AI in teaching, the challenges AI faces or will face in its application and educational activities, and the future direction of educational AI [8]. Every year, the development trend of modern technology application in education teaching is predicted, from which it is obvious that the application of AI and other technologies in education is driving unprecedented innovation and reform in education, in which AI shows the most powerful boost [9]. In addition, Tingzhou et al. have studied the impact of the application of AI technology on the learning process of students while proposing methods to optimize the factors that influence learning and make AI technology more relevant and practical to combine with education [10]. Tiffany Barnes, for example, analyzed and concluded that the application of AI technology can be an effective tool to promote learning and teaching efficiency through the application of AI in computer teaching. Rahim et al. explain what benefits the Internet can bring to teaching and learning [11]. It is mentioned that Internet applications not only help to improve the richness of teaching content but also help students to break through the time and space constraints and make better use of fragmented time for learning. Of course, Internet applications can also help teachers improve their teaching theories, understand students' learning needs, and ensure teaching quality.

The Knewton Adaptive Learning Platform not only collects data from students' online learning but also accurately analyzes and predicts students' proficiency in theoretical knowledge and practical processes and further analyzes students' interests and expectations and recommends appropriate content for continued learning based on the data collected and analyzed [12]. Some researchers have analyzed Purdue University's "Curriculum Signals" project, which has shown that the project can influence the process of educators' teaching activities and learners' learning process, use learning analytics to help educators understand the real learning situation of learners, and propose strategies and measures to improve teaching and learning in terms of learning patterns. MOOC Buddy, a teaching robot serving

the MOOC platform, can collect learners' relevant learning data and analyze learners' characteristics [13]. Through a series of effective analyses and feedback, the teaching-learning relationship between teachers and students can be made closer, and the teaching effect can be better presented while the students' learning efficiency and learning quality can be improved to a greater extent to realize personalized education. Pitić et al. explain that distance learning may become an indispensable part of future education and teaching in the context of the 5G era [14]. The authors make a more profound study of the online and offline as well as distance learning models through the situation reflected by the online teaching during this epidemic. Some of the current problems of distance learning are analyzed, and the prospects for development are projected.

Therefore, this paper starts from the perspective of the study of human subjectivity in the system of "AI + education", which is not only a reflection on the real problems highlighted by the application of AI in the new era from the philosophical aspect but also a further examination of the issue of human subjectivity in the new context of the times, which is a specific aspect of the issue of human subjectivity. In-depth study, which is conducive to the further development of artificial intelligence and the further strengthening of human subjectivity, is increasingly dazzling vibrancy.

### **3. 5G Joint Artificial Intelligence Technology in University English Education Innovation and Reform Application Design**

*3.1. 5G Joint Artificial Intelligence Technology Design.* AI technology [15, 16] has always been a technology that has been pursued and studied in-depth in this era, and the birth of 5G communication technology has had a profound impact on AI technology, which has been able to develop rapidly with the support of 5G technology, which is characterized by its ubiquitous network and the Internet of Everything, which can have a profound and positive impact on AI technology. This impact can be seen everywhere around us now, making it possible for us to get the first glimpse of AI technology. The city we live in is constantly being updated because of the rapid development of AI technology, whether it is cars, streetlights, cameras, or utility access hole covers [17]. 5G + AI technology is still very much in practice in the teaching field. Teachers can understand students' learning psychology and status promptly. The combination of teachers and AI technology can primarily help teachers to make effective classroom designs based on data analysis. It further promotes the implementation of the education idea of teaching according to students' abilities. We can develop different learning plans for different students, study different teaching strategies, and set up and distribute different teaching contents according to different students' characteristics as much as possible.

Second, AI is also very helpful for teachers to improve their teaching level. In the absence of AI, teachers can only change their teaching strategies by their own teaching expe-

rience or the problems reflected by some students. Often, teachers do not know whether students are adapting to their lecture style or whether students have mastered what they are doing to teach. In the traditional form, teachers usually verify the appropriateness of their methods through regular testing, as well as to understand how well students have mastered the knowledge points. The use of artificial intelligence can help teachers adjust their teaching methods, enrich their knowledge base, and understand the students' acceptance level, making education more humane and personalized. With the rapid development of 5G, teaching, as an important window for Chinese culture dissemination, should be more innovative in education and teaching methods. For example, VR/AR can be incorporated into the classroom to give students an immersive learning experience; through distance learning, explore the teaching mode of dual-teacher classroom and master classroom to balance teaching resources; take advantage of big data analysis to collect learning data and customize personalized teaching programs according to students' characteristics; give full play to the advantages of sharing network resources and continuously expand the existing teaching materials to realize teaching materials sharing. When our Chinese language teaching workers and teachers of Confucius Institutes teach abroad with new teaching methods, they not only show the world our world-leading technology level but also can effectively improve the quality of Chinese language teaching, as shown in Figure 1.

Second, online teaching teachers can choose to record or live-stream their lessons. Some teachers upload the video footage of the class to the platform in advance of the class to increase the time for oral practice in the class. Both the video of the class and the teaching material can be saved for a long time so that students can watch it anytime and anywhere. If students do not hear something in class or do not understand it well, they can study it carefully after class [18]. This approach breaks the time and space constraints and helps students to grasp the knowledge points better. Some of the online teaching software nowadays is equipped with an automatic test paper correction function, and the corrected data will be analyzed twice and pushed to the teacher. The teacher can test the students' learning results in the classroom and produce the results in the classroom, which can help students to focus on the class. It not only ensures teaching effectiveness but also improves teachers' teaching efficiency. According to the system's data analysis, teachers can personalize after-class assignments and conduct targeted teaching. Nowadays, some online teaching software has the function of automatic correction of test papers, and the corrected data will be analyzed twice and pushed to teachers. Teachers can test students' learning results in the classroom and produce grades in the classroom, which can help students focus on their lessons. It not only ensures teaching effectiveness but also improves teachers' teaching efficiency. Based on the system's data analysis, teachers can personalize after-class assignments and conduct targeted teaching.

The Butterworth filter is suitable for use as a bandpass filter because it has a smooth frequency response curve in

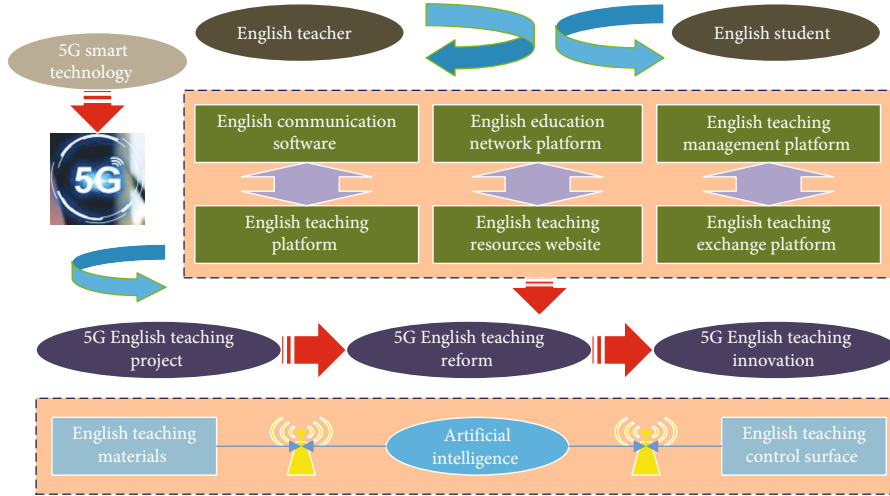


FIGURE 1: 5G college English education joint artificial intelligence.

the passband, and the system function is chosen as the Butterworth function, whose amplitude squared function is expressed in the following equation.

$$|H(j\omega)|^2 = \frac{1}{1 - \varepsilon^2 (j\Omega/j\Omega_c)^{2N}}, \quad (1)$$

$$N \leq \ln \left( \frac{\lambda}{s} \right). \quad (2)$$

The classical wavelet denoising method also has the disadvantage that a better denoising effect can be obtained by estimating the noise variance for noisy signals, but this is not the case when the noise is small. Because the background noise in sEMG is generally considered to satisfy the Gaussian model. Therefore, in this paper, the estimation of the noise variance of EMG based on the Gaussian mixture model (GMM) is used, and the minimum Gaussian coefficient is used as the estimate of the noise variance.

$$p(y|\theta) = \sum_{m=1}^k a_m p(y|\theta_m). \quad (3)$$

The preprocessed sEMG cannot be directly sent to the pattern classifier for action classification, and the useful information in the sEMG that can represent different action categories, i.e., features, must be extracted, and then a variety of different features are used to form a feature vector. Classical feature extraction is based on the characteristics of the data using some method to represent a pile of data with some value, which can be regarded as a transformation of the data from high to low dimensions to some extent and has a certain effect on dimensionality reduction, which helps to carry out the pattern recognition smoothly later. As an integral part of the pattern recognition process, feature extraction plays a crucial role in the effectiveness of recognition. Each gesture has its unique features, and the features that can reflect the general similarity of the same gesture and the difference between different gestures are called good

features. SEMG feature extraction is to extract the good features from the surface EMG signal, and the good features extracted will directly affect the final gesture recognition result.

$$\text{MAV} = \frac{1}{N} \sum_{i=1}^N x_i, \quad (4)$$

$$\text{VAR} = \frac{1}{N+1} \sum_{i=1}^N x_i^2. \quad (5)$$

Because of the random nature of sEMG, there is instability in the time domain analysis. The nature of muscle tissue activity can be reflected in the frequency domain analysis by first performing a Fourier transform on the signal to obtain the spectral distribution and then analyzing the signal characteristics in the frequency domain. The frequency distribution of the original surface EMG signal after preprocessing is in the range of 20-500 Hz and contains a lot of information about muscle contraction. A total of four frequency domain features were selected for extraction in this paper.

$$\text{PKE} = \min (p_1, p_2, p_3, \dots, p_N), \quad (6)$$

$$\sum_{i=1}^{\text{MDF}} p_i = \frac{1}{N} \sum_{i=1}^N p_i. \quad (7)$$

In the field of pattern recognition, the selection and extraction of feature vectors have a crucial impact on the recognition effect of the whole system. The more dimensions of the feature vectors do not guarantee a better final classification effect. On the contrary, too many dimensions of the features may have some features containing only little classification information and duplicate classification information among the features, and the existence of these features will cause the pattern recognition system to work inefficiently or even ineffectively, as shown in Figure 2.

English is a special subject compared to other subjects, and students need to broaden their horizons and gain more



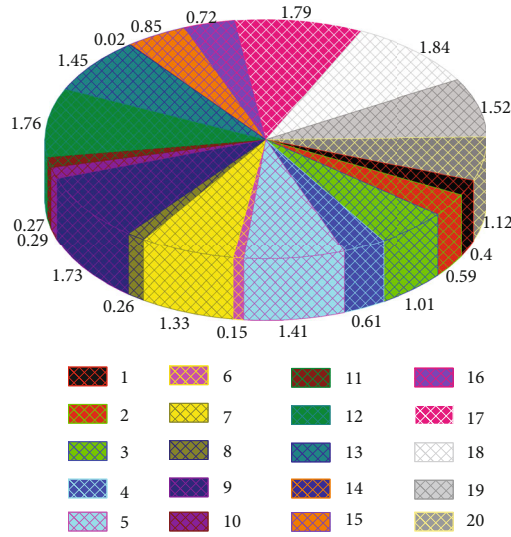


FIGURE 2: Relative importance of each feature.

rich contexts to have a better understanding of humanities and society abroad. Therefore, this requires that students' learning content not only exist in the classroom but also the students' vision is particularly important, and teachers need to use some conditions and resources to improve students' listening and speaking ability and comprehensive quality. Outside the classroom, more extracurricular activities should be organized to provide students with a free environment to learn English, and through continuous practice, students' interest in learning English can be naturally stimulated. However, the lack of a standard for measuring English listening and speaking ability has led many teachers to be ambiguous about the direction of listening and speaking ability cultivation and cultivation measures, external listening and speaking training does not achieve the purpose of ability enhancement, and there must be a complete, scientific, and reasonable evaluation standard [19]. This evaluation standard should be a process evaluation rather than a summative evaluation standard, which requires teachers to be able to intervene appropriately in the process of students' learning, to be able to solve the problems that constantly arise in learning, to praise and motivate students' performance promptly, and to gradually build up students' sense of achievement to better develop students' English listening and speaking ability. The cultivation of English subject listening and speaking ability is a long process, break the traditional teaching model, student-oriented, teacher-led way to carry out the combination of internal and external teaching activities, only in this way can make the junior high school students' English listening and speaking ability be improved.

To avoid these undesirable effects, this paper extracts and selects features based on reading a lot of related literature and calculates the relative importance of each feature value by random forest to prove that the selected features play an effective role in the final classification effect. The decision tree algorithm also performs sample partitioning based on the features while classifying, so the information gain-based decision tree algorithm also completes the com-

parison of the importance of the input set of features. The score of each feature in each decision tree is calculated after the training is completed, and finally, the total score of each feature in the random forest is calculated. The results are normalized so that the sum of the relative importance of the features is equal to 1.

$$G^m = 1 + \sum_{i=1}^N (p_i^m)^2. \quad (8)$$

To better promote teaching and learning reform, problems such as failure to synchronize hardware and software construction, emphasis on effects rather than effectiveness, large gaps between theoretical research and practical needs, the imperfect management system in the field of practice, inadequate funding to support educational technology, and copying foreign theories and experiences must be solved as soon as possible. The article concludes by predicting the future development trend of educational technology and argues that comprehensive research on computer education applications should be given top priority and combined with practical and supportive research on educational technology, joint research on the psychology of learning in the technological environment, and the design of learning activities, to further improve the construction of educational technology disciplines. This indicates that computer education application is an important direction for the future development of educational technology, and computer education applications will glow with continuous vitality.

*3.2. Analysis of the Application of Innovation and Reform in University English Education.* After the completion of the training base, the training room is now equipped with 12 sets of wired VR devices and 6 sets of wireless VR devices, which can meet the use and training needs of 36 students at the same time. Through the latest teaching platform, I load the cognitive knowledge and simulation disassembly and assembly training that needs to be operated in the classroom into the platform, so that the teacher can teach the students more quickly, quickly, and efficiently through the online simulation training board. At the same time, students can repeat the simulation training both in class and out of class [20]. I found that more than 80% of the instructors felt that virtual reality (VR) and augmented reality (AR) technology could help students understand the functions of various vehicle parts more intuitively while reducing the rate of damage to parts and injuries during hands-on training. More than 90% of the students said that they are now willing to take practical classes and learn the corresponding theoretical knowledge, compared with the previous teacher's boring explanation, now can be more intuitive understanding of knowledge. And compared to the previous face of the real equipment, now, students for the virtual hands-on training can be more bold hands-on operation. In the spare time of the class can also be repeated practice, more conducive to the improvement of their skills.

By building an AI teaching platform with a 5G transmission network as the carrier, I use edge computing technology

to link high-definition webcams to capture the demeanor, expressions, and movements of teachers and students in the classroom in real-time. Using AI artificial intelligence technology, the data is analyzed in the background and pushed to teachers and schools, and parents can also grasp the learning status of students in the classroom if necessary, as shown in Figure 3.

During independent learning, students can realize the ability to follow the system to read out the sentences or for the articles displayed in the system. Afterward, the system scores and the intelligent evaluation results are displayed on the screen in real-time [21]. The intelligent evaluation and real-time correction of students' pronunciation can help students to make corrections. The teacher sets the words, grammar, sentences, and articles needed for the lesson, and the students select the corresponding Chinese learning materials after entering the system, and the system leads the reading while the students learn and follow. After the learning is completed, the system will automatically generate a learning report. All students' reports will be summarized and sent to the teacher. The system also helps teachers to analyze and interpret students' data and help teachers to prepare and teach lessons in a differentiated way.

Holographic projection technology is a technology based on the principle of interference to record and display the three-dimensional image of an object, which is one of the 3D technologies. Currently, 3D movies are popular and people need to wear glasses to see the beautiful 3D images when watching 3D movies. But holographic projection technology is more than that; people can see the three-dimensional virtual world without wearing glasses, so holographic projection technology is also called virtual imaging technology. Its working principle is divided into two steps: the first step is based on the interference principle to record the light wave information of the object, i.e., to photograph the object; the second step is based on the diffraction principle to reproduce the light wave information of the object, i.e., to image and show the three-dimensional image of the object.

Speaking up in class is a psychological barrier for many students. They are afraid to speak up, afraid of making mistakes, and sometimes have some anxiety. When teachers design the classroom, they try to motivate students as much as possible, enliven the classroom atmosphere, stimulate students' interest in learning, and make them willing to speak up on their initiative [22]. This not only increases the fun of learning but also allows students to understand the linguistic context of idioms.

The classes are divided into beginner, intermediate, and advanced classes. Teachers develop different teaching methods for different levels of students. For the beginners' class, I try to use short and easy-to-understand vocabulary to teach and not to expand too much. This ensures that students have a good grasp of the basics. When teaching intermediate and advanced students, the content can be increased to focus on developing students' oral expression skills. If students are given enough time and space to learn at their own pace, most students will be able to effectively master what they have learned. Oral Chinese teaching should be

student-centered and develop students' Chinese communicative skills as the priority, as shown in Figure 4.

With online teaching becoming increasingly common, the transformation of teaching methods is an issue that cannot be ignored. Online teaching is not just a matter of transferring the offline classroom teaching model to online. The flipped classroom is a concept that is often mentioned nowadays. In the classroom, the teacher is student-centered and teaches through the actual needs of the students, with the students as the primary focus and the teacher as the secondary focus. In online teaching, teachers should give more trust to students and shift from the once fill-in-the-blank teaching to assisted teaching. Focus on developing students' independent learning skills so that they will learn and want to learn. Thus, students can develop good oral learning habits. To further explore what kind of relationship is presented between connotative elements and each influencing factor, the author used the same method to carry out the division of connotative elements and influencing factors after data transformation. Out ah D factors can be divided into internal and external factors of individuals, and I first performed correlation analysis on them separately.

## 4. Analysis of Results

*4.1. 5G Joint AI Algorithm Performance Results.* Different classifiers have their characteristics and apply to different features. To compare and select the most suitable classifier, this paper combines  $K$ -fold cross-validation to input the EMG feature values of the six gestures of the experimenter into each of the four classifiers introduced above and selects the classifier with better performance by comparing the correct rate of gesture recognition. The gestures recognized in this paper are six common gestures, and the sample size of each gesture for each experimenter is 500. Combined with fold-cross validation, the feature values of each gesture action are divided into five parts, one of the feature values is taken as the test set each time, the remaining four feature values are taken as the training set, and a total of five training classifications are performed. The correct rate of classification recognition is calculated based on the correct rate = number of correctly recognized samples/total number of tested samples. The statistics of gesture classification recognition results for experimenter 5 are shown in Figure 5.

The keywords that appeared more frequently were artificial intelligence, expert systems, distance education, machine learning, multimedia, knowledge engineering, computer-assisted instruction, and meaningful learning. It indicates that the applications of artificial intelligence in education include expert systems, machine learning, distance education, multimedia, and computer-assisted instruction. The development of artificial intelligence has provided new tools for educational research, along with a series of technologies such as speech recognition, pose recognition, expression recognition, and EEG recognition. To better analyze the learning process and the timing of educational interventions in-depth, a large amount of data analysis is necessary, and the use of visualization techniques in AI to describe knowledge

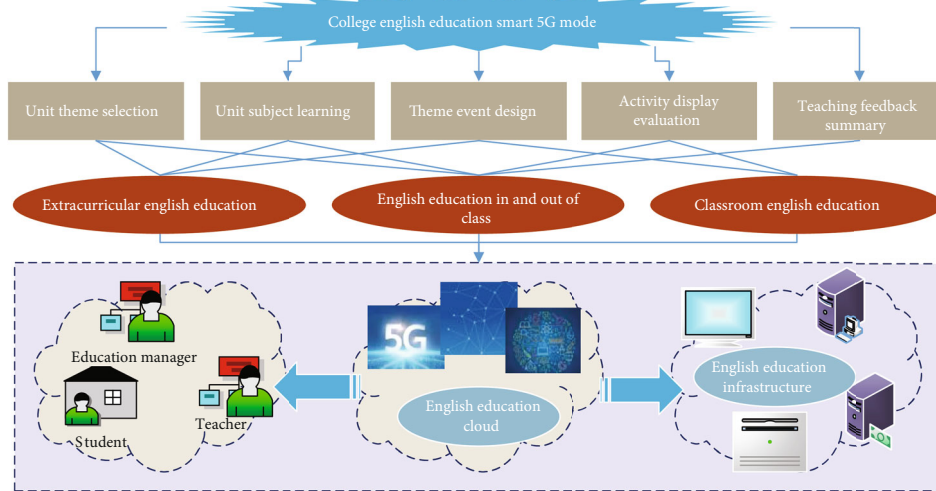


FIGURE 3: Network topology diagram.

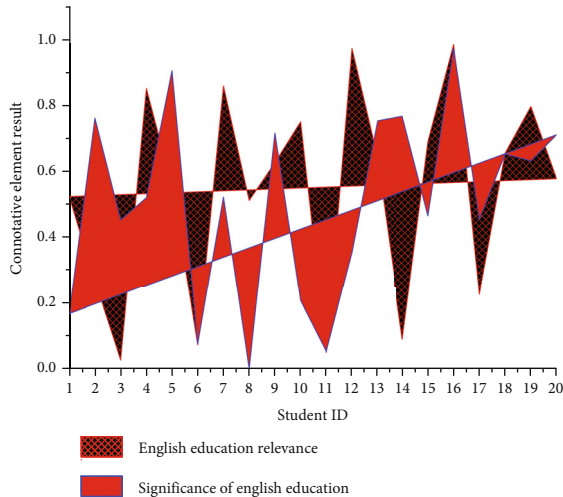


FIGURE 4: Correlation of inner elements.

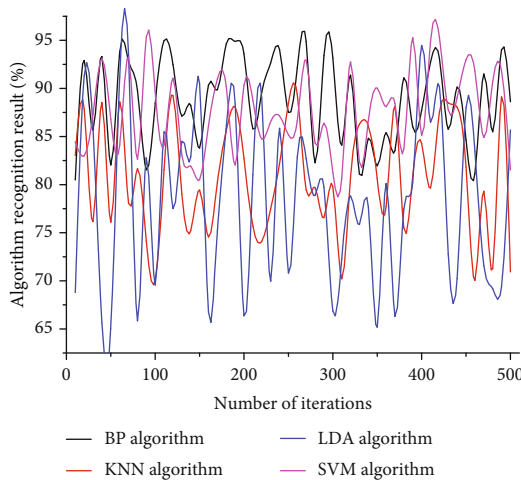


FIGURE 5: Algorithm performance comparison.

resources and their carriers, to mine, analyze, construct, map, and display knowledge and their interconnections is a very convenient way to do so. The maturity of these technologies has allowed AI to greatly increase the multidimensional experience of learning, thus, allowing learning to move toward deep learning supported by technology.

The successful research and development of intelligent teaching expert systems and intelligent teaching aid systems have injected new vitality into the development of artificial intelligence, and both have broad application prospects. With the development and application of neural network [28-29] and fuzzy technology [30] and the maturity of intelligent teaching expert system technology, people gradually realize that artificial intelligence applied to teaching is the need of education development. From this initially published article, we can learn that AI has great potential in the field of education, as intelligent teaching expert systems and intelligent teaching aid systems in expert systems, machine learning, and distance learning have gained rapid development in today's world, and they will continue to take advantage of the technology to build more realistic scenarios and experiences, allowing learners to better enter the situation and creatively problem solving, as shown in Figure 6.

The design of the NMA English speaking course objectives focuses on the specific design of specific colleges and universities, considering the current situation of international, local, and host universities. The design process should clarify the teaching objectives of teachers, the learning objectives of students, and the objectives of the course environment in the course. The course should give full play to the advantages of the university where it is located, the advantages of the location, and combine it with the needs of the students' course objectives. The NMA Oral English course that combines the specificity of the institution should be designed with specific course objectives for the needs of all levels. The content design of spoken English courses should focus on a combination of online resources, textbooks, and linguistic scholarship and should strive to provide the most

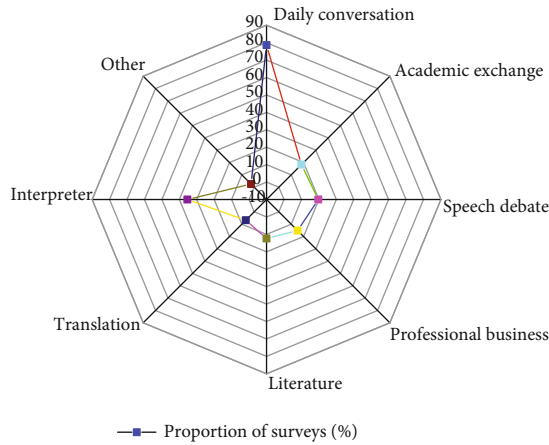


FIGURE 6: Analysis of the content needs of spoken English courses.

appropriate learning materials in situations where learners have the greatest need or are most receptive to knowledge.

**4.2. The Effect of Innovation and Reform Application in University English Education.** When the carrier of AI makes human subjectivity lose the balance between virtual and reality, the phenomenon of weakening human subjectivity gradually emerges from it. In the AI technology ecological environment, due to the unique openness and sharing of intelligent technology and intelligent environment, some information, speech, and interaction of people are supervised from all sides, but what needs attention is that even if they are supervised, people are more likely to make some bold, nonreal, uncritical, and spontaneous speech and ideas in the virtual world, which leads to the virtual self. The virtual self is cut off from the real self, making people living in the real world one-sided. People's emotional cognitive power is weakened, and people themselves are not able to reflect the objectivity of objective things dynamically, instead, people as subjects are also in a passive state. In such an environment, the slippery slope of human ethics as a subject will further lead to the weakening of human subjectivity. Artificial intelligence is the object of activity of human intelligent subjects for practical activities, and human influences restrict and change the object to make the object fit with the needs of the subject. The desubjectification of artificial intelligence is the subjectification of the object of artificial intelligence, and the desubjectification of artificial intelligence as an object is the counteraction of the human intelligent subject to control and transform the objective world caused by the practical activities that endanger the subject itself.

Continuously explore a new school governance model, encourage educational institutions to apply AI technology to carry out innovative changes in organizational structure and management system, establish a more efficient and low-cost operation mechanism, and promote campus management that is refined, personalized, and intelligent, so that school governance can reach a new level. Guide online open learning platforms to establish a learner-centered development concept, cultivate users' lifelong learning habits, provide users with rich and perfect personalized learning

resources, strengthen the supply capacity of education services, and cultivate more outstanding talents, as shown in Figure 7.

From the data in Figure 7, it can be concluded that in pair 1, the analysis of the data derived from the pretest and posttest of students' language perception ability can be seen, where  $t = -3.046$ , significance value  $F = 0.004 < 0.05$ , it indicates that there is a significant difference in the experiment, indicating that there is a significant change in the language perception ability of the students in the experimental class before and after the experiment; in pair 2, the analysis of the data is derived from the pretest and posttest of students' language comprehension ability. The analysis of the data derived from the pretest and posttest of students' language comprehension ability shows that where  $t = -3.010$  and the significance value  $F = 0.004 < 0.05$ , which means that there is a significant difference in the experiment, indicating that there is a significant change in the language comprehension ability of the students in the experimental class before and after the experiment; in pair 3, the analysis of the data derived from the pretest and posttest of students' language evaluation ability shows that where  $t = -1.011$  and the significance value  $F = 0.325 > 0.05$ , which means that there is no significant difference in the experiment, indicating that there is no significant difference in the language evaluation ability of the students in the experimental class before and after the experiment, although it has been improved.

To better analyze the change of students' interest in English listening and speaking, the options of the questions related to listening and speaking interest in the students' questionnaire were analyzed, as shown in Figure 8, the number of people who chose "very much in line" increased from 19.86% to 24.85% of the total number of students, with an increase of 5.01%, the number of people who chose "The number of people who chose" conform "increased from 30.16% to 38.07% of the total number of people, with an increase of 7.91%, and the number of people who chose" very much not conform decreased from 5.34% to 3.86% of the total number of people, with a decrease of 1.49%. This shows that students' interest in English listening and speaking is gradually increasing during the experimental teaching process, and it also reflects that students' English listening and speaking habits are also gradually changing.

There was an impact on students' performance when the English listening and speaking skill development strategy based on the intelligent speech system was carried out in the experimental class, the English performance of students in the experimental class improved, and there was a significant difference compared to the control class. The students' English listening and speaking skills improved in the experimental class. This shows that the strategy of developing the English listening and speaking ability of junior high school students based on an intelligent speech system is effective. Pay attention to the artistry of method and language in class organization. Teachers should be the organizers, guides, and facilitators of teaching activities. In the era of artificial intelligence, students will gradually transform from passive receivers of knowledge to explorers, discoverers, and collaborators. With the changes in the learning environment,

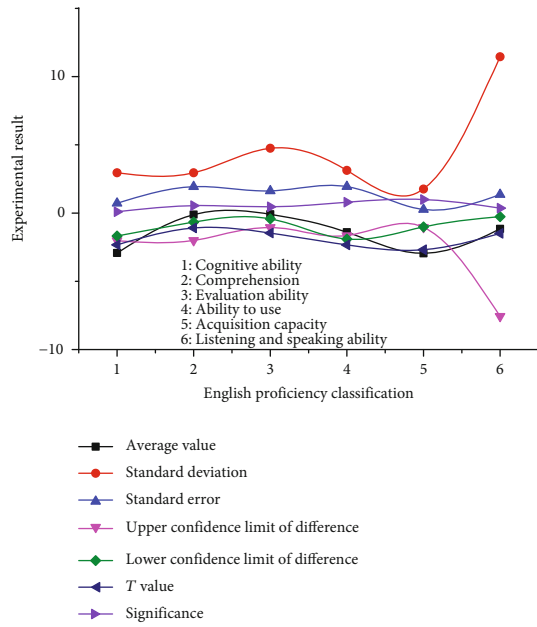


FIGURE 7: Paired sample test.

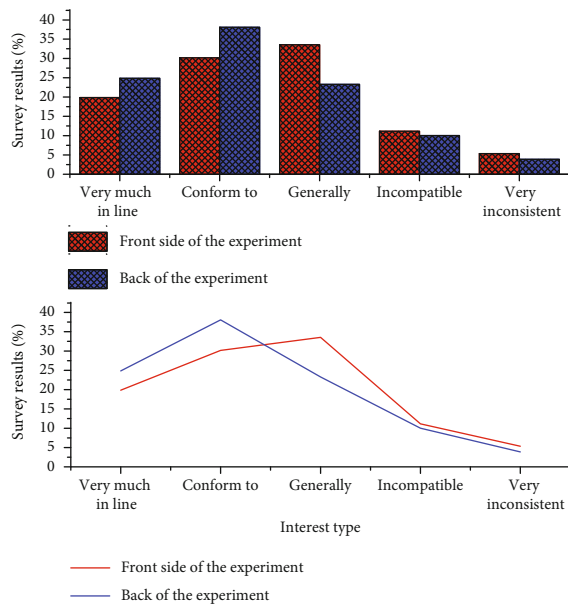


FIGURE 8: Changes in students' interest in listening and speaking English.

students must have the awareness of being in an intelligent age and the learning philosophy that conforms to this age and actively cultivate the literacy and interest in using artificial intelligence technology.

### 5. Conclusion

The advent of the era of artificial intelligence has triggered the work revolution, which inevitably leads to the education revolution. To actively respond to the work revolution and occupy the active posture, the education revolution must

be the first to start. The development and existence of human beings are not established and complete, but they are constantly striving for a free and perfect personality in the process of practice, and their subjectivity is constantly established and consolidated in the process of practice. The "Miseducation" system is a tool used in the field of education, and its original purpose of creation and development is to bring a better educational environment for human beings and to enable them to develop themselves comprehensively. The development and existence of human beings are not given and complete, but in the process of practice, they are constantly striving for a free and perfect personality, and in the process of practice, their subjectivity is constantly established and consolidated. The "Miseducation" system is a tool used in the field of education, and its original purpose of creation and development is to bring a better educational environment to human beings and enable them to develop comprehensively. I believe that the application in the context of 5G can bring a very good innovation for online oral teaching. It can not only greatly improve the teaching efficiency of teachers but also maximize the interest of students in learning. Therefore, this paper researches and discusses the online speaking teaching mode in the context of 5G, hoping to provide some new ideas and methods for teaching. The future is the era of collaboration between humans and intelligent machines. The application of artificial intelligence technology in English teaching will lead English teaching from closed to open and English learning from passive to active, which will completely change the time and space scenarios and supply levels of English education, making personalized and diversified education possible, and building A flexible, open, and lifelong personalized education ecosystem.

### Data Availability

The data used to support the findings of this study are included within the article.

### Conflicts of Interest

The author does not have any possible conflicts of interest.

### References

- [1] J. Zeng, "Artificial intelligence and China's authoritarian governance," *International Affairs*, vol. 96, no. 6, pp. 1441–1459, 2020.
- [2] C. Bartholomew, "China and 5G," *Issues in Science and Technology*, vol. 36, no. 2, pp. 50–57, 2020.
- [3] S. W. Harold and R. Kamijima-Tsunoda, "Winning the 5G race with China: a U.S.-Japan strategy to trip the competition, run faster, and put the fix in," *Asia Policy*, vol. 28, no. 3, pp. 75–103, 2021.
- [4] F. Sun, "Sino-UK science and technology collaboration in field of people's livelihood," *Journal of Chinese Economic and Business Studies*, vol. 18, no. 4, pp. 349–354, 2020.
- [5] H. Brands and C. Edel, "A grand strategy of democratic solidarity," *The Washington Quarterly*, vol. 44, no. 1, pp. 29–47, 2021.

- [6] A. Jash, "China's military-civil fusion strategy: building a strong nation with a strong military," *CLAWS Journal*, vol. 13, no. 2, pp. 42–62, 2020.
- [7] G. Austin, "The strategic implications of China's weak cyber defences," *Survival*, vol. 62, no. 5, pp. 119–138, 2020.
- [8] N. Noesselt, "A presidential signature initiative: Xiong'an and governance modernization under Xi Jinping," *Journal of Contemporary China*, vol. 29, no. 126, pp. 838–852, 2020.
- [9] P. Layton, "Artificial intelligence, big data and autonomous systems along the belt and road: towards private security companies with Chinese characteristics?," *Small Wars & Insurgencies*, vol. 31, no. 4, pp. 874–897, 2020.
- [10] L. Tingzhou, L. Sha, Z. Yuyang, and L. Wanying, "Policy analysis of educational services for the one belt one road initiative under the COVID-19 epidemic," *Beijing International Review of Education*, vol. 3, no. 1, pp. 72–91, 2021.
- [11] M. N. Rahim, "Post-pandemic of Covid-19 and the need for transforming education 5.0 in Afghanistan higher education," *Utamax: Journal of Ultimate Research and Trends in Education*, vol. 3, no. 1, pp. 29–39, 2021.
- [12] A. Pichetworakoon, N. Kooptarnond, and S. Ngamchuensuwan, "Economic and legal on the deploying of medical and healthcare robotics: case study on a comparison of the European Union (EU), South Africa, and Thailand. The journal of law, public administration and social science," *School of Law Chiang Rai Rajabhat University*, vol. 5, no. 2, pp. 21–43, 2021.
- [13] X. Wu, "Technology, power, and uncontrolled great power strategic competition between China and the United States," *China International Strategy Review*, vol. 2, no. 1, pp. 99–119, 2020.
- [14] G. Pitić, M. Kržić, A. Vuković, and M. Ilić, "Gaming industry in Serbia: a chance for a new industrial policy," *Ekonomika preduzeća*, vol. 68, no. 1-2, pp. 91–103, 2020.
- [15] W. Chu, P. S. Ho, and W. Li, "An adaptive machine learning method based on finite element analysis for ultra low-k chip package design," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, p. 1, 2021.
- [16] Z. Chu, M. Hu, and X. Chen, "Robotic grasp detection using a novel two-stage approach," *ASP Transactions on Internet of Things*, vol. 1, no. 1, pp. 19–29, 2021.
- [17] E. B. Kania and J. Costello, "Seizing the commanding heights: the PLA strategic support force in Chinese military power," *Journal of Strategic Studies*, vol. 44, no. 2, pp. 218–264, 2021.
- [18] C. S. Lai, Y. Jia, Z. Dong et al., "A review of technical standards for smart cities," *Clean Technologies*, vol. 2, no. 3, pp. 290–310, 2020.
- [19] R. Campa, "Fourth industrial revolution and emotional intelligence: a conceptual and scientometric analysis," *Changing Societies & Personalities*, vol. 4, no. 1, pp. 8–30, 2020.
- [20] Y. Xuetong, "Bipolar rivalry in the early digital age," *The Chinese Journal of International Politics*, vol. 13, no. 3, pp. 313–341, 2020.
- [21] C. Yang, S. Huan, and Y. Yang, "A practical teaching mode for colleges supported by artificial intelligence," *International Journal of Emerging Technologies in Learning (IJET)*, vol. 15, no. 17, pp. 195–206, 2020.
- [22] D. Wardianto, J. Jama, and S. Syahril, "The effectiveness of problem-project based learning to improve students' understanding toward gasoline motor," *International Journal of Scientific & Technology Research*, vol. 9, no. 3, pp. 4900–4902, 2020.

## Research Article

# Research on Modeling and Scheduling Methods of an Intelligent Manufacturing System Based on Deep Learning

Xiaoyi Lan<sup>1</sup> and Hua Chen<sup>2</sup> 

<sup>1</sup>School of Economic Management, Xi'an Technological University, Xi'an 710021, China

<sup>2</sup>School of Mechatronic Engineering, Xi'an Technological University, Xi'an 710021, China

Correspondence should be addressed to Hua Chen; [chenhua2035@126.com](mailto:chenhua2035@126.com)

Received 1 August 2021; Revised 15 August 2021; Accepted 16 August 2021; Published 22 September 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Xiaoyi Lan and Hua Chen. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Under the background of intelligent manufacturing, the modeling and scheduling of an intelligent manufacturing system driven by big data have attracted increasing attention from all walks of life. Deep learning can find more hidden knowledge in the process of feature extraction of the hierarchical structure and has good data adaptability in domain adaptation. From the perspective of the manufacturing system, intelligent scheduling is irreplaceable in intelligent production when the manufacturing quantity of workpieces is small or products are constantly changing. This paper expounds the outstanding advantages of deep learning in intelligent manufacturing system modeling, which provides an effective way and powerful tool for intelligent manufacturing system design, performance analysis, and running status monitoring and provides a clear direction for selecting, designing, or implementing the deep learning architecture in the field of intelligent manufacturing system modeling and scheduling. The scheduling of the intelligent manufacturing system should integrate intelligent scheduling of part processing and intelligent planning of product assembly, which is suitable for intelligent scheduling of any kind and quantity of products and resources.

## 1. Introduction

Over the years, people have conducted extensive research on the application technology of artificial intelligence in a manufacturing system. As a result, a variety of expert systems and knowledge-based systems for specific fields have been developed, forming a series of “intelligent arc islands” [1]. With the deepening of research, people gradually realize that the future manufacturing automation should be intelligent integration and the further improvement of manufacturing automation depends on the self-organization ability of the whole manufacturing system [2]. The intelligent manufacturing system can be regarded as a complex system composed of various intelligent subsystems or intelligent nodes. Each subsystem completes the distributed collaborative solution on the basis of material, energy, and information exchange and converts manufacturing system resources into products [3]. The purpose of the intelligent manufacturing system is to simulate the skills and expert knowledge of the manufacturing technology by integrating knowledge engineering and

manufacturing software system, robot vision, and robot control, so that intelligent machines can produce without manual intervention [4]. The purpose of intelligent scheduling is to complete the production tasks of production processing and machine assembly on the premise of meeting various constraints of the manufacturing system. In the process of intelligent production, due to the frequent changes of production objectives, such as the acceleration of product upgrading, the intelligent manufacturing system can also be adjusted according to the new production tasks to ensure that the machine can complete the production tasks with high utilization [5].

The purpose of the intelligent manufacturing system is to turn human intellectual activities into intelligent activities of manufacturing machines. The physical basis of the intelligent manufacturing system is an intelligent machine, which includes intelligent machining machine tools with various programs, tool and material conveying devices, detection and test devices, assembly devices, etc. [6]. Intelligent automation of product processing and manufacturing has

become the main direction today. It is particularly important to make job shop scheduling intelligent, which is an urgent need for the survival and development of enterprises [7]. In the intelligent manufacturing process, production is fully automated and unmanned. Once unreasonable processing process arrangement occurs, the machine may collide and cause safety accidents [8]. From the perspective of the manufacturing system, intelligent scheduling is irreplaceable in intelligent production when the number of workpieces is small or the products are changing [9]. The intelligent dispatching system is an important part of IMS. Intelligent scheduling should be a scheduling technology that comprehensively studies the intelligent scheduling of part processing and intelligent planning of product assembly from the overall index. It must be highly universal, intelligent, and flexible [10]. This paper expounds the outstanding advantages of deep learning [11–14] in intelligent manufacturing system modeling [15], provides an effective way and powerful tool for intelligent manufacturing system design, performance analysis, and operation state monitoring, and provides a clear direction for intelligent manufacturing system modeling and scheduling field selection, design, or implementation of deep learning architecture.

IMS refers to a highly flexible and integrated method to simulate the intelligent activities of human experts in the manufacturing process through computers in all links of the manufacturing process, that is, to analyze, judge, reason, conceive, and make decisions on intelligent problems in manufacturing [16]. With the emergence of new manufacturing modes such as computer-integrated manufacturing, flexible manufacturing system, fine production, and virtual manufacturing, it has indeed brought great convenience to enterprises in terms of manufacturing cost and product quality assurance [17]. Intelligent scheduling is the key to ensure the efficient, reliable, and safe production of the intelligent manufacturing system, so the research on intelligent scheduling has very important practical significance. The planning, design, scheduling, control, operation status monitoring, intelligent manufacturing system modeling, system maintenance, and performance analysis of the intelligent manufacturing system must be supported by corresponding models [18]. Deep learning is also called the deep neural network (DNN). It adopts the hierarchical structure of multiple neural layers and extracts information from input data through layer-by-layer processing. This “deep” layer structure allows it to learn the representation of complex raw data with multiple levels of abstraction. From the original input, DNN automatically discovers complex structures in large datasets and learns useful features layer by layer. DNN has been widely used in visual recognition and language understanding because of the characteristics of feature learning, and this feature learning ability has become its key advantage [19]. The neural network [20–22] architecture aims at modeling the high-level representation of data and classifying or predicting patterns by stacking multilayer information processing modules in the hierarchy [23]. The intelligent machining unit is an important intelligent node of the intelligent manufacturing system. In-depth research on modeling and scheduling of the intelligent manufacturing system based

on the deep learning neural network can improve the intelligence level of the intelligent machining unit.

## 2. Related Work

In recent years, artificial intelligence algorithms [24–26] have gradually entered the stage of solving industrial manufacturing scheduling problems. Literature [27] proposes a simple optimization algorithm based on pareto by constructing a mathematical model of multiobjective flexible industrial manufacturing scheduling based on parallel machines. Literature [28] established a new hybrid genetic algorithm, which solved the problems of premature, slow searching speed and poor convergence effect on the manufacturing scheduling problem of operation industry. Literature [29] starts from the perspective of multivariety and small-batch production, adopts standard datasets to prove and solve the manufacturing scheduling of the flexible operation industry, considers the factors such as the maximum completion time, machine load rate, and total load rate as multiple scheduling objectives, and uses the method of analysis and deconstruction to verify the effectiveness and practicability of the algorithm. Literature [30] solves the flowshop problem of two machines, in which the optimization object is the minimum processing time and the dynamic programming algorithm is mainly used. Literature [31] conducted a detailed study on the flow line problem from three aspects of system modeling, solution method, and algorithm performance evaluation in 2020 and proposed the use of the Lagrangian algorithm to solve the static scheduling problem of flow line operations. Literature [32] used the branch and bound method to solve the multimachine flowshop-type problem.

Traditional scheduling methods can still achieve satisfactory results for simple and small-scale scheduling problems, but it is difficult to solve the actual scheduling problems that are more complex, large scale, and difficult to model with ordinary methods. Literature [33] puts forward an encoding method based on EDD-LPT when using genetic algorithms to solve scheduling problems to solve various unexpected situations encountered during scheduling, combining various real examples of manufacturing in the remanufacturing workshop, and constructing. Many different scheduling models have been developed to provide theoretical and technical guarantees. Literature [34] puts forward the concept of autonomous and open-system machine tools in the future environment and believes that in the future advanced manufacturing system, processing machine tools should be autonomous and open. As an experimental prototype, they developed an intelligent machine tool based on a vertical machining center. Literature [35] proposed a method of controlling system beacons to avoid system deadlock, but when manufacturing system considerations gradually become more complex, deep learning models will explode and bring huge trouble to beacon control. Literature [36] proposes to avoid the deadlock of the system by controlling the strict minimum beacon of the system, that is, to maintain the activity of the system by adding a controller to each minimum beacon. This article combines deep learning to



optimize the modeling and scheduling of intelligent manufacturing systems, provides an effective way for the design, performance analysis, and monitoring of operating conditions of intelligent manufacturing systems, and selects, designs, or implements deep learning architectures for intelligent manufacturing system modeling and scheduling.

### 3. Materials and Methods

*3.1. Structural Model.* The intelligent manufacturing system is a digital, networked, and intelligent manufacturing based on ubiquitous state awareness in virtual reality. In the process of product design, it involves the knowledge or professional knowledge of various fields related to the product and also involves the comprehensive processing and utilization of these multifield knowledge, experience, and data. There are many difficulties in the process of transforming the applicability of scientific and technological achievements in manufacturing small and microenterprises. From the perspective of system modeling, the manufacturing system is divided into resource elements and process elements and resource elements are divided into equipment resource elements and information resource elements. The former includes various machine tools and tools, while the latter includes various CAD, CAPP, and other software that provide accompanying information such as design data, process flow, and machining data for process elements [37]. As one of the intelligent nodes of the intelligent manufacturing system, the intelligent processing unit is also composed of many intelligent nodes. It includes the flexible manufacturing system, flexible assembly system, intelligent fault detection system, and human experts. The structural model of the manufacturing system mainly describes the attributes, quantities, and interrelationships of equipment resource elements and process elements.

Figure 1 shows the mode structure of a completely distributed intelligent manufacturing system, which is composed of many relatively autonomous self-agents connected by a system bus. Self-agents, which are relatively autonomous, are connected to the communication network in the form of intelligent nodes and accomplish tasks together through the collaboration and cooperation of all nodes. In implementation, modules with strong cohesion and low coupling with other tasks can be separated to form collaborative tasks. This can reduce the boundary constraints and facilitate the early start of the next task. In the process of project execution, the workflow provides a process monitoring tool, which is convenient for managers at all levels to know the work progress and status. This can relatively reduce the difficulty of the task and speed up the completion of the task. The collaborative design system for concurrent engineering should provide enough facilities to form a collaborative environment, so that collaborative participants can use the facilities in the environment to describe collaborative requirements. Because the concurrent design of concurrent engineering is divided into intergroup and intragroup collaboration, the work of members in the group is closely related and collaboration generally requires synchronization. However, the design work between groups may need to be

coordinated at intervals, so the collaboration can be carried out asynchronously. Intelligent nodes can include any form of intelligent units, and other intelligent nodes can be connected to the system bus, such as intelligent CAPP (computer-aided process planning) nodes and intelligent CAD (computer-aided design) nodes. Each intelligent node is an independent self-agent.

The system organization mode of the intelligent processing unit is a distributed mode, which adopts the centralized control and scheduling mode in its local implementation, and is managed by a core node, which is responsible for the dynamic allocation of tasks and resources, coordinates the competition and cooperation among other nodes through global planning, and acts as an arbitrator to resolve conflicts. Other nodes apply for tasks from core nodes in a certain way according to their own capabilities. The nodes receiving tasks can complete tasks independently, and at the same time, they can request cooperation from other nodes through core nodes. Under the macrocontrol of core nodes, they can establish temporary cooperative relations with other cooperative nodes to complete tasks together.

*3.2. Behavior Model.* With the development of information technology and manufacturing technology, the complexity of products is increasing, the amount of information to be processed in product manufacturing is increasing, and the division of social specialties is becoming more and more detailed, which makes the cooperation between enterprises increasingly close. Multilevel supplier collaboration aims at the realization process of complex products, major equipment, or large-scale projects, and with the help of various information and management technologies and means, enterprise resources distributed in different spaces and times and belonging to different partners are quickly and effectively organized into a unified organism. In the manufacturing process of complex products, from the initial conceptual design to the mass production of products, suppliers are involved and undertake the development and production tasks of important parts. In the allocation process, the priority of task allocation between internal and external resources is considered based on the allocation strategy and the optimal supplier of the product collaborative development chain is selected by integrating the factors of task influence and supplier influence.

The intelligent manufacturing system is a very complex large-scale system, which is a multifactor, high-order, and nonlinear system, and the traditional modeling method is very difficult. It is characterized by discreteness, difficulty in on-line detection, uncertainty of the process model, rapidity of process, and instability of processing multilevel information feedback [38]. The intelligent modeling method can be realized by fuzzy mathematics, neural network, and other methods. Scheduling problem is actually a constrained optimization problem. In the past research on scheduling methods, the assumption of constraints was often limited to resource constraints. The change of the manufacturing process state is determined by the change of the resource element and process element state. Equipment resource elements have three valid states: idle state, busy state, and

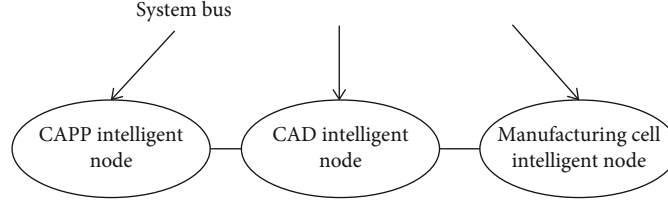


FIGURE 1: Distributed intelligent manufacturing system structure.

fault state. Process elements show different states according to their different processing stages.

Artificial intelligence technology can very accurately monitor the stability of electronic information systems during information transmission to ensure the safety and accuracy of information input or output. Only in this way can information processing be carried out better and a good auxiliary role be played for the implementation of the intelligent manufacturing system. Using the existing technology to establish an artificial intelligence manufacturing environment, the intelligent design operation process is shown in Figure 2.

Generally speaking, the parts in the machining process can be in different stages such as clamping, machining, transportation, cleaning, and measurement. Each process may go through the processes of waiting, loading, processing, unloading, finishing, etc. However, due to the different process routes of the parts, the change paths between these states are also different. The scheduling of intelligent processing units in the intelligent manufacturing system should integrate intelligent scheduling of part processing and intelligent planning of product assembly, which is suitable for intelligent scheduling of any kind and quantity of products and resources. This method can automatically make intelligent decisions according to product requirements and system resources, so as to maximize the autonomy of intelligent processing units.

**3.3. Control Decision Model.** Because the utilization of equipment resources by various activities in the manufacturing process is random in time and the processing tasks are also dynamic, there is competition in the utilization of equipment resource elements and choice in process elements. This needs to be solved by system scheduling, and the scheduling must establish the control decision model of the system. For the intelligent scheduling of intelligent processing units, besides the system resource constraints, the priority order constraints of each working procedure and the priority order constraints of part completion should also be considered [39]. These priority constraints are provided by the assembly planning system of the flexible assembly system and meet the requirements of optimal assembly. The distributed networked IMS prototype system is the comprehensive embodiment of the research achievements of this project in the basic theory and technology of intelligent manufacturing. The basic idea is to start from the essential characteristics of the intelligent manufacturing system and, in the distributed manufacturing network environment, according to the basic idea of distributed integration, apply the theory and method of the multiagent system in distributed artificial intelligence, and focus on the flexible and intelligent integration of the

manufacturing unit and network-based manufacturing system. The structure of the deep neural network in this article is shown in Figure 3.

The determination of the weight of each index in the technical analysis of intelligent manufacturing is a very important step in the calculation process of the technical evaluation index. Since the weights of each process in the entire intelligent manufacturing process are not the same, it is necessary to use the analytic hierarchy process to determine the specific weights, in which each index is calculated as follows:

(1) Disassembly index

$$\lambda_1 = \frac{m_{\text{ideal}} \times t_{\text{ideal diss}}}{t_{\text{actual diss}}}, \quad (1)$$

where  $m_{\text{ideal}}$  is the ideal number of parts,  $t_{\text{ideal diss}}$  is the ideal disassembly time, and  $t_{\text{actual diss}}$  is the actual disassembly time

(2) Cleaning index

$$\lambda_2 = \frac{m_{\text{ideal}} \times 1}{l_{\text{cleaning}}}, \quad (2)$$

where  $l_{\text{cleaning}}$  is the cleaning score and  $m_{\text{ideal}} \times 1$  is the ideal cleaning score. There are four main cleaning methods: blowing, wiping, baking, and washing. When different cleaning methods are used, the cleaning scores are 1, 3, 3, and 6, respectively

(3) Check index

$$\lambda_3 = \frac{m_{\text{ideal insp}}}{m - m_{\text{replaced}}}, \quad (3)$$

where  $m_{\text{ideal insp}}$  is the ideal number of inspections,  $m$  is the number of parts, and  $m_{\text{replaced}}$  is the number of replacements

(4) Detection index

$$\lambda_4 = \frac{m_{\text{test}} \times t_{\text{ideal test}}}{t_{\text{actual test}}}, \quad (4)$$

where  $m_{\text{test}}$  is the number of parts to be inspected,  $t_{\text{ideal test}}$  is the ideal inspection time, and  $t_{\text{actual test}}$  is the actual inspection time

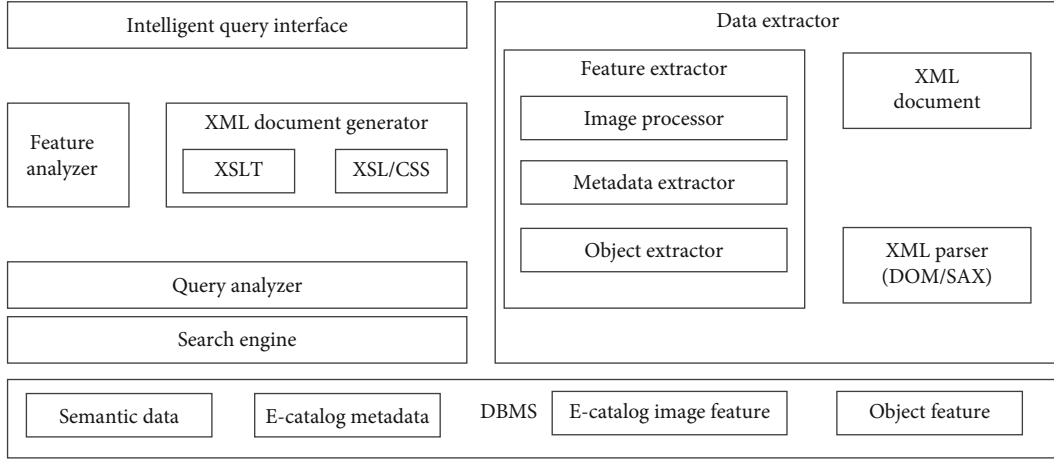


FIGURE 2: Manufacturing intelligent design operation process.

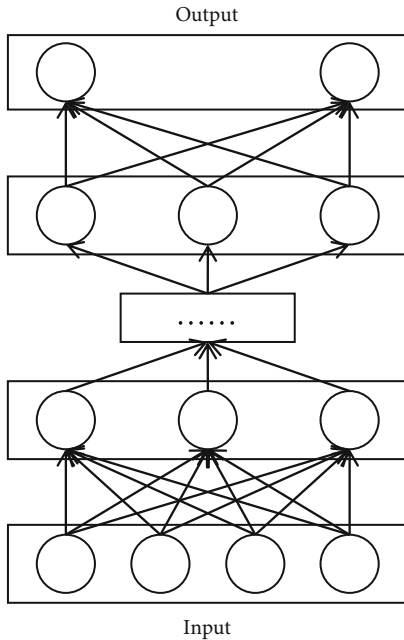


FIGURE 3: Deep neural network learning model.

(5) Assembly index

$$\lambda_5 = \frac{m_{\text{ideal}} \times t_{\text{idealass}}}{t_{\text{actualass}}}, \quad (5)$$

where  $m_{\text{ideal}}$  is the ideal number of parts,  $t_{\text{idealass}}$  is the ideal assembly time, and  $t_{\text{actualass}}$  is the actual assembly time. After determining the calculation of each index, the technical evaluation index  $\lambda_T$  can be further calculated:

$$\lambda_T = \left( \sum_j \frac{W_j}{\lambda_j} \right)^{-1}, \quad (6)$$

where  $j = 1, 2, \dots, 5$ ,  $\lambda_j$  is the aforementioned 5 indexes, and  $W_j$  is the weight of the 5 indexes. The range of the calculated

$\lambda_T$  is  $[0, 1]$ . When  $\lambda_T > 0.5$ , it shows that it is suitable for intelligent manufacturing from a technical point of view.

The scheduling method of the intelligent processing unit should be based on the overall index and comprehensively study the scheduling technology integrating intelligent scheduling of part processing and intelligent planning of product assembly. It must be highly universal, intelligent, and flexible. Each activity first requests the response of the process element server and then requests the response of the corresponding resource server according to the attributes of the selected parts.

#### 4. Scheduling Analysis of the Intelligent Manufacturing System Based on Deep Learning

*4.1. Reconfiguration and Optimization of Cell Resources in Intelligent Manufacturing.* The purpose of data acquisition is to collect relevant data such as vibration, acoustic emission, current, velocity, temperature, and other signals from various sensor sources deployed on the equipment system. Data processing mainly includes data cleaning and other strategies for preprocessing. Through the signal processing technology and dimension reduction strategy, fault sensitive features are extracted and selected from the original signals and data fusion is carried out on multisensor signals. The purpose of intelligent manufacturing system modeling is to extract the production factors in the system and express their relationship reasonably, which can be used to analyze the performance or existing problems of the system. The complexity of the complex system lies in its complex structure, many variables, strong coupling among system variables, high nonlinearity, time delay, and time variability. There are many objects to realize health management, and the status signals containing health information are complex and diverse, as well as dynamic, static, continuous, and discrete, which usually requires a combination of modeling and forecasting techniques of intelligent manufacturing systems.

In engineering design, not only rich professional knowledge and basic design data but also practical experience and knowledge of experts are needed. Through the analysis and prediction of data, the traditional way of operation judgment

based on intuition and fuzzy judgment will be gradually replaced. The relationship between the number of nodes and processing time is shown in Figure 4.

Assume the intelligent processing unit shown in Figure 1 as follows: there are  $n$  parts; the  $i$ th part has  $n_i$  processes; there are  $m$ -type resources, and the number of  $s$ th-type resources is  $r_s$ ; the completion time of each part has priority constraints. The completion time of each part has priority constraints; the total completion time of all parts of the system is required to be the shortest, and the completion time of each part is recorded as  $x_i$ .

For the convenience of description, the following symbols are introduced:  $n$  is the number of parts;  $n_i$  is the number of processes of the part;  $m$  is the number of resource types;  $r_s$  is the number of resources of each type;  $R_i$  is the process pair  $[j, l]$  set of the  $i$ th part  $p$ . Among them, process  $j$  has priority over process  $l$ ;  $Q_i$  is the process pair  $[j, l]$  set of the  $i$ th part  $p_i$ , where processes  $j, l$  have no priority constraint relationship;  $I_i$  is the process set that can be arranged for the first time;  $N_{sq}$  is the use of the  $q$ th process set of  $s$  resources;  $P$  is the set of parts with priority completion time constraints;  $t_{il}$  is the processing time of process  $l$  of the  $i$ th part  $p_i$ ;  $x_{ij}$  is the completion time of process  $j$  of the  $i$ th part  $p_i$ . The completion time is the last process of each part in the scheduling sequence.  $i = 1, 2, \dots, n$ ;  $j = 1, 2, \dots, n_i$ .  $x_{ij}$  should satisfy the following inequalities:

$$\begin{aligned}
x_{ij} - x_{il} + t_{il} &\leq 0, \quad \text{for all } [j, l] \in R_i, \\
x_{ij} - x_{il} + t_{il} &\leq 0 \text{ or } x_{il} - x_{ij} + t_{ij} \leq 0, \quad \text{for all } [j, l] \in Q_i, \\
x_{ij} - x_{il} + t_{ij} &\leq 0 \text{ or } x_{ij} - x_{il} + t_{il} \leq 0, \quad \text{for all } [j, l] \in N_{sq}, \\
x_{ij} - x_{il} + t_{il} &\leq 0, \quad \text{for all } [j, l] \in R_i, \\
x_k - x_i &\leq 0, \quad \text{for all } [K, j] \in P, \\
t_{ij} - x_{ij} &\leq 0, \quad i = 1, 2, \dots, n; j = 1, 2, \dots, m.
\end{aligned} \tag{7}$$

In the given neural network model, if it represents a neuron, the energy function of the network is

$$\begin{aligned}
E = & \sum_{i=1}^n x_i + \sum_{i=1}^n \sum_j \sum_l H_1 \cdot F_1(x_{ij} - x_{il} + t_{il}) \\
& + \sum_{i=1}^n \sum_j \sum_l H_2 \cdot [\min(F_1(x_{ij} - x_{jl} + t_{jl}), F_1(x_{jl} - x_{ij} + t_{ij}))] \\
& + \sum_{s=1}^m \sum_{q=1}^{r_s} \sum_j \sum_l H_3 \cdot [\min(F_1(x_{ij} - x_{il} + t_{jl}), F_1(x_{jl} - x_{ij} + t_{ij}))] \\
& + \sum_k \sum_i H_4 \cdot F_1(x_k - x_i) + \sum_{i=1}^n \sum_{j=1}^{n_i} H_5 \cdot F_1(t_{ij} - x_{ij}).
\end{aligned} \tag{8}$$

Among them,  $F_1(\cdot)$  is the penalty function;  $H_1, H_2, H_3, H_4, H_5$  is a sufficiently large normal number. Deep neural networks can converge to a stable state, which corresponds

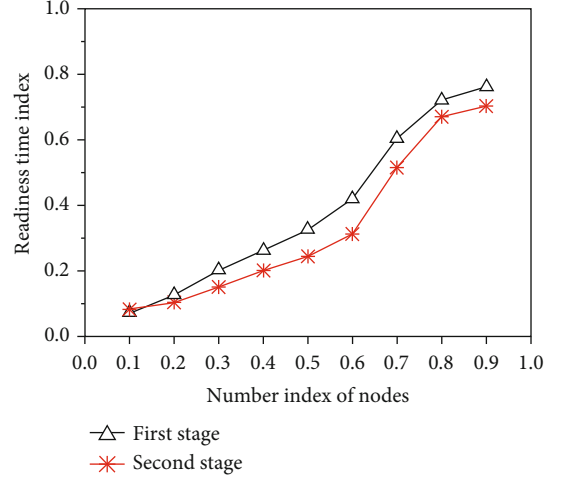


FIGURE 4: Number of nodes and processing time.

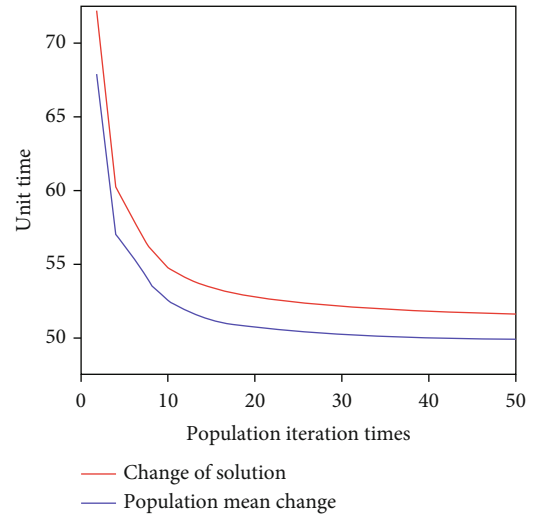


FIGURE 5: The population mean and the change of the optimal solution.

to an optimal solution of the problem. The population mean and the optimal solution changes are shown in Figure 5.

The main task of scheduling is to allocate the limited resources of the system reasonably and optimize some performance indexes. These resources can be machine tools, tools, fixtures, etc. System equipment status data is precious wealth, which contains abundant health information, and is the cornerstone of the data-driven intelligent manufacturing system modeling and prediction technology. Too many sensors will cause extra burden to health management objects, and too many sensor data will also cause data redundancy, which will bring difficulties to feature extraction and fault analysis.

**4.2. Dynamic Characteristics of NC Machine Tools in the Intelligent Manufacturing System.** Because of the complexity and uncertainty of the actual system, it is difficult to find a clear mathematical model. The method based on the failure physical model depends on the failure physics of the key components of the system, such as wear, fatigue, and aging.

TABLE 1: Characteristic value corresponding to each characteristic.

Features	A1	A2	A3	A4	A5	A6	A7	A8	A9
Characteristic value	75.5	88.7	87.8	95.6	74.5	76.8	87.3	92.3	75.4

The failure physics of different components has different failure processes and laws. It is costly and time-consuming to master the wear, fatigue, and aging process laws of key components. The performance of various intelligent devices in the intelligent manufacturing system plays an important role in the normal operation of the intelligent manufacturing system. With the improvement of automation, informationization, and intelligence of the machining system, the dynamic characteristics of modern CNC machine tools are also required to be improved accordingly. Whether the knowledge-driven method can model and predict an efficient intelligent manufacturing system depends on whether there is a complete expert system knowledge base, and it is difficult to deal with new faults without matching related rules. In the process of running, the expert system cannot get new knowledge from reasoning examples and it is very difficult to solve some novel faults and some marginal problems of the system design.

The configuration of components and the selection of performance parameters of CNC machine tools will directly affect the dynamic characteristics of the system. For the whole computer control system, starting with the electromechanical system composed of a servomotor and mechanical device, the mathematical model is established by using the laws of the mechanical system and electrical system. The transfer function of the mechanical system is

$$G_1(s) = \frac{C(s)}{\theta_m(s)} = \frac{L}{2\pi} nK_s \frac{1}{J_e s^2 + B_e s + K_s} = \frac{K_0 \omega_n^2}{s^2 + 2\xi \omega_n^2 + \omega_n^2}, \quad (9)$$

where:

$$\begin{aligned} K_0 &= \frac{Ln}{2\pi}, \\ \omega_n &= \sqrt{\frac{K_s}{J_e}}, \\ \xi &= \frac{B_e}{2} \sqrt{\frac{1}{J_e K_s}}. \end{aligned} \quad (10)$$

The transfer function of the electrical system is

$$\begin{aligned} G_2(s) &= \frac{\theta_m(s)}{V_a(s)} \\ &= \frac{K_l/R_a}{(T_e s + 1)[(J_1 + J_m)s^2 + B_m s + K_m(1-n)] + K_e K_l s/R_a}, \end{aligned} \quad (11)$$

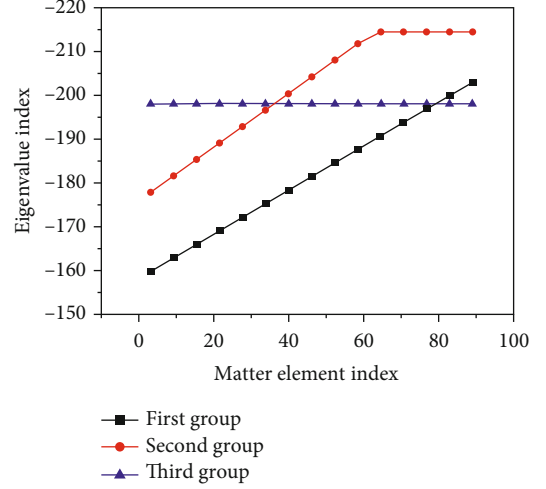


FIGURE 6: Relationship between eigenvalues and matter element.

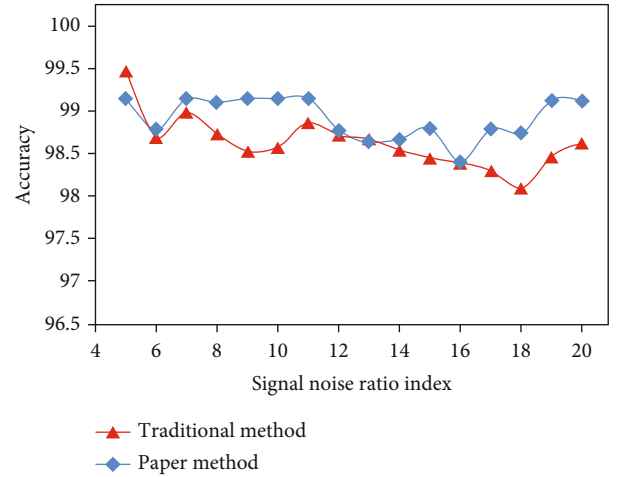


FIGURE 7: Modeling effectiveness of different methods.

where

$$T_e = \frac{L_a}{R_a}. \quad (12)$$

For matter elements, A is the evaluation characteristic, that is, the index. Some eigenvalues can be calculated by the system, and the corresponding eigenvalue of each feature is shown in Table 1. Figure 6 is the relationship between eigenvalues and matter elements in three groups of experimental data.

System modeling and simulation is an important means to analyze and design the system. It can easily put forward measures to improve the existing system and predict its dynamic performance and modify its parameters before

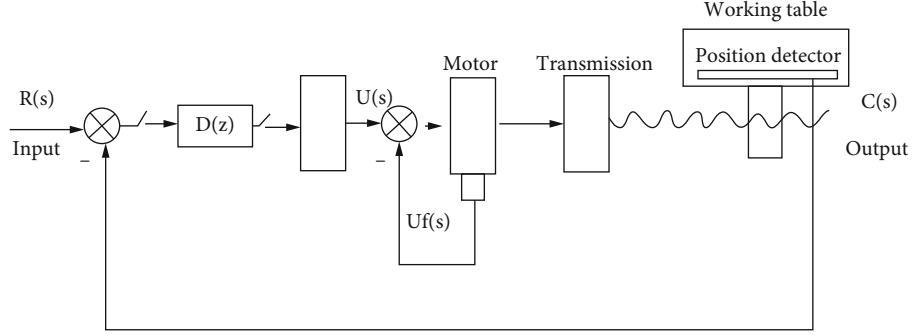


FIGURE 8: CNC machine tool closed-loop control system.

the system is established, so as to achieve the optimal design effect. The modeling effectiveness of different methods is shown in Figure 7.

In addition to the electromechanical control system of the CNC machine tool closed-loop control system, there is also a digital computer  $D(z)$  that completes digital signal processing, a sampler that completes analog-to-digital signal conversion, and a zero-order holder that completes digital-to-analog signal conversion. So the structure of the whole system is shown in Figure 8.

In which,

$$G(s) = \frac{K_v G_1(s) G_2(s)}{1 + K_v K_f S G_2(s)}. \quad (13)$$

$G(s)$  is the transfer function of the controlled object.

Ignoring the effect of the interference signal  $n(t)$ , that is,  $N(S) = 0$ , the output response  $C_R(Z)$  of the system to the input  $R(Z)$  can be obtained:

$$C_R(z) = \frac{D(z)G'(z)}{1 + D(z)G'(z)} R(z), \quad (14)$$

where

$$G'(z) = Z \left[ \frac{1 - e^{-Ts}}{s} G(s) \right]. \quad (15)$$

Equipment status data contains abundant information on the equipment status, including the relationship between equipment components and random errors, which can truly describe the system performance. In practical engineering, it is uneconomical or even impossible to obtain mathematical models and it is difficult to express expert knowledge and experience effectively. In this case, data-based technology is more used. By comparing the scheduling results of the general modeling method and deep neural network, as shown in Figure 9. It can be seen that the modeling and scheduling effect of the intelligent manufacturing system based on the deep neural network is better than that of general modeling methods.

The intelligent manufacturing system must realize typical intelligent signs of state perception, real-time analysis, independent decision making, and precise execution and real-

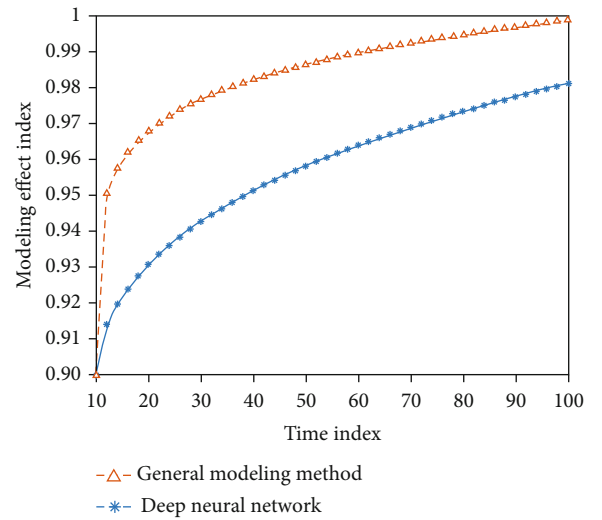


FIGURE 9: Comparison of the deep neural network and the general modeling method.

ize visual monitoring [40, 41] of the product manufacturing process. For modeling and simulation of a computer control system, the general method is to establish a mathematical model describing the movement of the system according to the dynamic relationship among various physical quantities in the system, then to model the mathematical model twice and turn it into a simulation model, and to finally send the simulation model to a computer for the solution. With the improvement of the computer processor performance and the reduction of its cost, artificial intelligence technologies such as the expert system, fuzzy system, and neural network have gradually developed. Based on these technologies, there is an increasing interest in developing artificial intelligence solutions in the field of equipment health management. Compared with the traditional diagnosis method, the diagnosis based on intelligent technology can improve the performance with the least manual help and can be easily expanded and modified and can be adjusted according to the new data.

## 5. Discussion

Compared with traditional machine learning methods, due to the deep architecture and the learning method of learning useful features layer by layer, the deep learning model can

automatically find the fault mode features contained in large datasets, so as to realize automatic feature learning without a special feature extractor. In some application scenarios, the fault diagnosis and prediction accuracy of the support vector machine and traditional artificial neural network can reach the accuracy equivalent to that of the deep learning model, but in terms of robustness in a noisy environment, the deep learning model is better than the former. The traditional artificial neural network is easy to fall into the trap of gradient explosion and overfitting. The deep learning model can effectively avoid similar problems. However, in terms of computational speed, because the computational complexity of the deep learning model is higher than that of the support vector machine and traditional artificial neural network, the training time is more than that of the latter.

The deep learning model is essentially developed from a traditional neural network. Improve the model generalization ability, strengthen the local feature learning, optimize the model parameters using optimization algorithm, and improve the data preprocessing technology, multimodal data fusion, and mixed integration of multiple models, and other technical routes are continuously integrated into the deep learning framework, which effectively improves the model generalization ability, multimodal data fusion, and mixed integration of multiple models. With the introduction of various technologies, the computational complexity of the model will inevitably increase. How to extend the fault prediction and diagnosis model trained by historical data in order to ensure the accuracy of medium- and long-term fault prediction and diagnosis is a major challenge. In order to improve the generalization ability of the model, the transfer learning technology has been introduced to construct different domain-adaptive deep learning models and simulate different working conditions on the training data as much as possible. However, for practical application, the generalization ability of the model still needs to be strengthened. In general, the training of the deep learning model needs massive data. However, it is not easy to obtain the health information of the actual complex equipment system. Due to historical reasons, there are often few or no sample data, while the running equipment is in normal working state for most of the running time and a small amount of fault sample data can be collected.

## 6. Conclusions

Under the global market competition environment, manufacturing globalization is an inevitable trend of manufacturing development and it is also an opportunity and challenge faced by Chinese manufacturing enterprises. With the arrival of the customized manufacturing trend, the future development of the manufacturing industry will become more and more intelligent, which puts forward higher and higher requirements for the scheduling of intelligent manufacturing systems. Traditional manual scheduling methods can no longer meet the requirements of intelligent manufacturing systems and complex manufacturing systems. The modern manufacturing industry is facing the requirements of individuation and diversification and the phenomenon of a shorter and shorter product cycle,

which requires the product design and process design to be coordinated. The scheduling of intelligent processing units in the intelligent manufacturing system should integrate intelligent scheduling of part processing and intelligent planning of product assembly, which is suitable for intelligent scheduling of any kind and quantity of products and resources. The foundation of intelligent manufacturing is digital manufacturing. By constructing the digital platform of the intelligent manufacturing system, the digital, networked, and intelligent system architecture operation model of the intelligent manufacturing system is built.

Compared with the traditional diagnosis method, the diagnosis based on intelligent technology can improve the performance with the least manual help and can be easily expanded and modified and can be adjusted according to the new data. In the process of manufacturing system modeling, through the modeling and simulation analysis of basic manufacturing cells, the bottom-up modeling is realized and the basic manufacturing cells are continuously expanded upwards, and finally, the manufacturing system model is established. In the aspect of scheduling algorithm research, the deep neural network proposed in this paper has solved the multiobjective scheduling problem but it needs further research on the determination of the convergence algebra and coding mode. By establishing standardized models for various types of scheduling problems, the research on scheduling algorithms will be more targeted, which will help to speed up the research on intelligent scheduling algorithms.

## Data Availability

The data used to support the findings of this study are included within the article.

## Conflicts of Interest

The authors do not have any possible conflicts of interest.

## Acknowledgments

This study was supported without any funding.

## References

- [1] L. Guang and X. Yang, "Research on industrial big data application combined with deep learning," *Big Data*, vol. 4, no. 5, pp. 6–17, 2018.
- [2] C. Zhiqiang C. Xudong et al., "Application of deep learning in equipment failure prediction and health management," *Chinese Journal of Scientific Instrument*, vol. 40, no. 9, pp. 209–229, 2019.
- [3] L. Yong, L. Weidong, and H. Liang, "On the application of deep learning in artificial intelligence courses," *Educational Research*, vol. 3, no. 10, pp. 105–106, 2020.
- [4] D. Xiangjun and M. Wang, "The network fine-tuning learning process of automatic recognition of ground-based cloud images by migration deep learning," *Electronic Devices*, vol. 43, no. 6, pp. 71–75, 2020.

- [5] M. Zarandi, A. Khadangi, F. Karimi, and I. B. Turksen, "A computer-aided type-II fuzzy image processing for diagnosis of meniscus tear," *Journal of Digital Imaging*, vol. 29, no. 6, pp. 677–695, 2016.
- [6] H. Boudhar, M. Dahane, and N. Rezg, "New dynamic heuristic for the optimization of opportunities to use new and remanufactured spare part in stochastic degradation context," *Journal of Intelligent Manufacturing*, vol. 28, no. 2, pp. 1–18, 2017.
- [7] A. M. Farid, "Measures of reconfigurability and its key characteristics in intelligent manufacturing systems," *Journal of Intelligent Manufacturing*, vol. 28, no. 1, pp. 1–17, 2017.
- [8] S. Tan, X. Dong, and C. Weidong, "Design of intelligent auxiliary system for crane operation based on speech recognition," *Manufacturing Automation*, vol. 42, no. 6, pp. 8–10, 2020, 26.
- [9] Y. Zhang, A. Bernard, R. K. Gupta, and R. Harik, "Feature based building orientation optimization for additive manufacturing," *Rapid Prototyping Journal*, vol. 22, no. 2, pp. 358–376, 2016.
- [10] H. Tengxiao, T. Dunbing, Z. Haitao et al., "Research on real-time dynamic scheduling mechanism of intelligent manufacturing system based on agent," *Machinery Manufacturing and Automation*, vol. 6, pp. 164–168, 2017.
- [11] M. Zhao, A. Jha, Q. Liu et al., "Faster mean-shift: GPU-accelerated clustering for cosine embedding-based cell segmentation and tracking," *Medical Image Analysis*, vol. 71, article 102048, 2021.
- [12] S. Qi, X. Ning, G. Yang et al., "Review of multi-view 3D object recognition methods based on deep learning," *Displays*, vol. 69, article 102053, 2021.
- [13] M. Zhao, Q. Liu, A. Jha et al., "VoxelEmbed: 3D instance segmentation and tracking with voxel embedding based deep learning," 2021, <https://arxiv.org/abs/2106.11480>.
- [14] Y. Gu, A. Chen, X. Zhang, C. Fan, K. Li, and J. Shen, "Deep learning based cell classification in imaging flow cytometer," *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 2, pp. 18–27, 2021.
- [15] Z. Luo, "Application and development of electronic computers in aero engine design and manufacture," *ASP Transactions on Computers*, vol. 1, no. 1, pp. 6–11, 2021.
- [16] Y. K. Liu, X. S. Zhang, L. Zhang, F. Tao, and L. H. Wang, "A multi-agent architecture for the scheduling of platform-based intelligent manufacturing systems (English)," *Frontiers of Information Technology & Electronic Engineering*, vol. 20, no. 11, pp. 14–42, 2019.
- [17] J. Stergar and U. Maver, "Review of aerogel-based materials in biomedical applications," *Journal of Sol-Gel Science and Technology*, vol. 77, no. 3, pp. 738–752, 2016.
- [18] N. Nahas, "Buffer allocation and preventive maintenance optimization in unreliable production lines," *Journal of Intelligent Manufacturing*, vol. 28, no. 1, pp. 85–93, 2017.
- [19] L. Zhengyang, Y. Xin, Y. Yang et al., "Intelligent manufacturing in orbit space: distributed scheduling modeling and optimization," *System Engineering Theory and Practice*, vol. 39, no. 3, pp. 705–724, 2019.
- [20] W. Cai, Z. Wei, Y. Song, M. Li, and X. Yang, "Residual-capsule networks with threshold convolution for segmentation of wheat plantation rows in UAV images," *Multimedia Tools and Applications*, pp. 1–17, 2021.
- [21] M. Li, G. Zhou, W. Cai et al., "MRDA-MGFSNet: network based on a multi-rate dilated attention mechanism and multi-granularity feature sharer for image-based butterflies fine-grained classification," *Symmetry*, vol. 13, no. 8, p. 1351, 2021.
- [22] W. Chu, P. S. Ho, and W. Li, "An adaptive machine learning method based on finite element analysis for ultra low-k chip package design," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, 2021.
- [23] H. Zhuangzhi, Y. Wang, and C. Jianfei, "Evaluation method of industrial alliance manufacturing resource allocation in intelligent manufacturing environment," *Machine Design and Research*, vol. 35, no. 1, pp. 161–164, 2019.
- [24] W. Cai, Z. Wei, R. Liu, Y. Zhuang, Y. Wang, and X. Ning, "Remote sensing image recognition based on multi-attention residual fusion networks," *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 1, pp. 1–8, 2021.
- [25] R. Liu, X. Ning, W. Cai, and G. Li, "Multiscale dense cross-attention mechanism with covariance pooling for hyperspectral image scene classification," *Mobile Information Systems*, vol. 2021, Article ID 9962057, 15 pages, 2021.
- [26] C. Yan, G. Pang, X. Bai et al., "Beyond triplet loss: person re-identification with fine-grained difference-aware pairwise loss," *IEEE Transactions on Multimedia*, p. 1, 2021.
- [27] L. Zhifeng, C. Wei, C. Yang et al., "Cloud platform for intelligent manufacturing workshop scheduling based on digital twins," *Computer Integrated Manufacturing System*, vol. 25, no. 6, pp. 1444–1453, 2019.
- [28] S. M. Weiss, A. Dhurandhar, R. J. Baseman et al., "Continuous prediction of manufacturing performance throughout the production lifecycle," *Journal of Intelligent Manufacturing*, vol. 27, no. 4, pp. 751–763, 2016.
- [29] S. Torabi, S. Alibabaei, B. B. Bonab, M. H. Sadeghi, and G. Faraji, "Design and optimization of turbine blade preform forging using RSM and NSGA II," *Journal of Intelligent Manufacturing*, vol. 28, no. 6, pp. 1409–1419, 2017.
- [30] S. Hanoun, A. Bhatti, D. Creighton, S. Nahavandi, P. Crothers, and C. G. Esparza, "Target coverage in camera networks for manufacturing workplaces," *Journal of Intelligent Manufacturing*, vol. 27, no. 6, pp. 1221–1235, 2016.
- [31] C. Xianzhao, "Multi-objective workshop scheduling analysis of intelligent manufacturing system," *Science and Technology Wind*, vol. 406, no. 2, pp. 34–34, 2020.
- [32] Z. Yaqin, W. Junliang, B. Jinsong et al., "Research on general data model for intelligent control of knitting production," *China Mechanical Engineering*, vol. 30, no. 2, pp. 143–148+219, 2019.
- [33] X. Wu, T. Songling, S. Zheng, and X. Zhongxia, "Real-time scheduling method of intelligent manufacturing workshop for open structure products," *Computer Integrated Manufacturing System*, vol. 270, no. 10, pp. 105–115, 2020.
- [34] B. O. Gombé, G. G. Mérou, K. Breschi et al., "A SAW wireless sensor network platform for industrial predictive maintenance," *Journal of Intelligent Manufacturing*, vol. 30, no. 4, pp. 1617–1628, 2019.
- [35] C. Dong, L. Yongxian, Z. Jia et al., "Research on vehicle visual inspection method based on deep learning algorithm," *Manufacturing Automation*, vol. 41, no. 3, pp. 119–122+125, 2019.
- [36] Y. Jing, F. Zhu, and W. Yang, "Research on multi-variety and small-batch production scheduling system based on on-time delivery," *Machinery*, vol. 44, no. 4, pp. 10–14, 2017.
- [37] O. Huabing, "Research status and development trend of intelligent manufacturing technology," *Journal of Shanghai Dianji University*, vol. 21, no. 6, pp. 14–20+27, 2018.



- [38] L. Zhicun, J. Yang, Y. Menghong et al., "Research and application of intelligent dispatching system for environmental protection stockyard in iron and steel enterprises," *Metallurgical Automation*, vol. 42, no. 6, pp. 18–23, 2018, 53.
- [39] L. Shengli, "Design of Tanggang stainless steel steelmaking logistics tracking and intelligent dispatching management system," *Electronics World*, vol. 540, no. 6, pp. 204–204, 2018.
- [40] J. Zhang, Y. Liu, H. Liu, and J. Wang, "Learning local-global multiple correlation filters for robust visual tracking with Kalman filter redetection," *Sensors*, vol. 21, no. 4, p. 1129, 2021.
- [41] J. Zhang, J. Sun, J. Wang, and X. G. Yue, "Visual object tracking based on residual network and cascaded correlation filters," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–14, 2020.

## Research Article

# Artificial Intelligence System for College Students' Physical Fitness and Health Management Based on Physical Measurement Big Data

Li Ai <sup>1,2</sup>

<sup>1</sup>College of Physical Education, University of Sanya, Sanya 572022, China

<sup>2</sup>College of General Education, Guangxi University of Technology (Preparatory), Guigang 537100, China

Correspondence should be addressed to Li Ai; [hxciaili@126.com](mailto:hxciaili@126.com)

Received 3 August 2021; Revised 24 August 2021; Accepted 25 August 2021; Published 21 September 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Li Ai. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Most of the current health management products are used in medical institutions and generally do not pay enough attention to the student population. Based on this, this paper designs a student-oriented and functional autonomous health management system. This paper proposes a personal health management system based on a multidimensional data model based on the main social characteristics of the population with chronic diseases and the actual needs of personal health management for chronic diseases. The value of various health data for health management is deeply analyzed and mined, and a multidimensional model data warehouse is constructed according to relevant national health data standards to create a standard data platform for intelligent health warning and disease risk assessment. This paper researches and designs a closed-loop personal health management method based on the Plan-Do-Check-Action (PDCA) cycle management model, with detailed functional design in four aspects: health data collection and recording, health assessment, health planning, and tracking and execution. This paper researches health data collection, processing, and storage technologies and adopts HDFS data storage technology, html, css, Java Script, java, and other software development technologies, combined with j Query, UEditor, Date Range Picker, and other plug-ins, as well as SMS email generation interface, wireless Bluetooth transmission interface, etc. This system web and mobile application platforms are designed and developed. Relational database is used as the system database, and a snowflake-type multidimensional data model is designed. Finally, the functions and performance of this system were tested, and the development and trial run of the basic version have been completed.

## 1. Introduction

As population continues to age, people's lifestyles and environments become more complex, leading to a very large population of chronic diseases, increasingly serious sub-health problems, and rising cancer incidence and mortality [1]. The 2018 joint study on medical reform released by the National Health and Family Planning Commission states that nearly 300 million people suffer from chronic diseases. Chronic diseases have become the main cause of death among residents [2]. Statistics show that deaths from chronic diseases account for 79.5% and 85.3% of the total deaths in rural and urban areas in China, respectively. The main sources of death from chronic diseases are cardiovas-

cular diseases, cancer, and slow respiratory diseases, among which the proportion of hypertension, diabetes, and slow respiratory diseases are 25.2%, 9.7%, and 9.9%, respectively, which shows that the harm of chronic diseases is serious and long-term [3]. With the increasing improvement of people's living and cultural standards, people are also paying more and more attention to personal health management, gradually shifting the focus of personal health management from illness and then medical treatment to prevention-based continuous long-term self-health management [4]. According to the World Health Organization statistics, the main factors leading to chronic diseases and their proportion are lifestyle 60%, genetic 15%, social 10%, medical 8%, and climatic 7%. Therefore, a good lifestyle can reduce the prevalence of

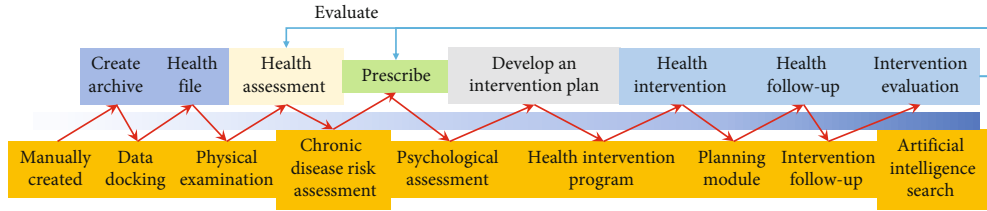


FIGURE 1: Health management process.

chronic diseases or improve the worsening trend of existing chronic diseases.

To develop a good lifestyle, people need comprehensive and accurate information about their health, long-term tracking of their physical condition, and timely intervention to correct poor lifestyles, of which comprehensive and accurate health information is the most important part [5]. The lack of transparency and asymmetry of medical information makes it difficult for individuals to participate in medical and health management, and it is easy to generate excessive medical care and waste medical resources [6]. As mentioned above, if people want to participate in the whole process of their own health management continuously in the long term, the temporary data obtained from medical visits under the traditional model is far enough to meet people's demand for health management. With the rapid advancement of computer technology, the gradual popularization of home medical devices, and the introduction of national health China 2030 and other information policies, the "Internet + health care" health service model has been developed rapidly, and personal health management systems have been created [7]. Personal health management system can be understood as a computer application that can prevent and control the occurrence and development of physical diseases, reduce medical costs, improve the quality of life and self-management awareness, improve lifestyle-related health risk factors, and make the long-term continuous process of personal health management in a safe and confidential environment. Health information collection, health testing, health assessment, health planning, and health intervention are the main means to continuously improve health status. The purpose is to prevent the occurrence and development of chronic noncommunicable diseases, improve the quality of life, and reduce medical costs [8]. The personal health management system is a perfect combination of advanced information technology and medical technology, mainly highlighting the individual as the center; the system assists individual users to carry out self-health management, so that health management can be carried out scientifically and on the basis [9]. It mainly helps healthy people and subhealthy people to establish an orderly and healthy lifestyle by maintaining health and promoting health (as shown in Figure 1). By collecting and recording personal health information, the system then cleans, stores, and analyzes user's health data, which can ensure the accuracy and integrity of the health data throughout the life cycle, making user's data information more accurate, continuous, and valuable, thus ensuring the accuracy of system's health assessment and early warning of user's health condition, achieving the goals of reduc-

ing the risk state, preventing the occurrence of diseases, and improving life. This will reduce the risk status, prevent diseases, improve the quality of life, save medical resources, and reduce the cost of medical treatment for patients [10]. At the same time, the development of personal health management system is of great importance in improving the efficiency and quality of medical services, changing the medical model, and realizing scientific cost control and actuarial management. There are tens of thousands of personal health management application platforms in China, but most of them provide a single health management model, only providing the management of one or several health data [11]. There are also many health management application platforms that mainly focus on medical consumption and cooperate with medical institutions to provide online registration for medical treatment, etc., without providing precise health management modules to users [12]. These single-mode or paid health management platforms fail to create a complete health management mechanism for users, and the vast majority of existing health management systems fail to enter rural households, staying only for users with higher living standards.

This project is based on a student-oriented approach to implement a health management system based on the complete process of health management. The system is designed to assist students in the complete process of detecting, evaluating, and guiding themselves on potential risk factors for their health. It is a professional system belonging to the medical field. The purpose of the system is to motivate students to implement autonomous health management, i.e., students improve their ability to manage their own health through multiple ways such as self-learning, self-building, and self-establishing databases and rely on their own strength to monitor and care for their health indicators in real time, which is a dynamic autonomous health management model. From the student level, the use of health management system can reduce the expenditure of medical expenses; most students are provided with good health maintenance and protection services, then the risk factors of diseases can be controlled within a certain range without clinical reactions. The use of health management systems can reduce students' health risk factors, making it a virtuous circle in which students become more and more aware of their health management, and independent health management makes students secure.

## 2. Related Work

As early as 1929, world's first health management organization was established in Los Angeles, USA, which proposed a

comprehensive approach to health management by focusing on prevention to reduce the incidence of disease and conducted practical exploration of health management. In the early 1970s, the U.S. government formulated rules and recommendations for health maintenance and established and passed a health maintenance bill for this purpose. In the following years, the bill was implemented by establishing health research centers and developing health assessment [13]. In the following years, the act was implemented by establishing health research centers, developing health assessment tools, and so on. In 2010, the U.S. government launched a Decade of Health program to improve national health and disease prevention awareness [14]. Japan's health care system features private, one-on-one, long-term health follow-up assessments [15]. This tracking and assessment service is family-based, with a comprehensive health record that is updated and maintained over time, and recommendations for assessment are made accordingly.

The business of health management has been very mature in foreign countries, and with the expansion of the scope of the industry and changes in the composition of the disease, many countries have invested in the management of chronic diseases and other modern health persistent disease research, through sensors, the Internet, and other technologies to establish a sound personalized chronic disease monitoring system [16]. The concept of health management emerged in the United States and other Western countries in the 1990s and was first introduced by the American insurance industry. Today, the United States is still the global leader in healthcare development, especially in the field of health management [17]. The closed-loop health management system mainly targets people with subhealth or subdisease, through the steps of health data collection, health assessment, plan making, guidance intervention, and user feedback, cultivating users' good living habits, improving their self-health management awareness and self-healing power, and achieving the effect of controlling the occurrence or deterioration of diseases at the root. In addition to the United States, health management is also developing rapidly in developed countries such as the United Kingdom, Japan, and Korea. With the development of Internet communication technology, many health management applications have emerged with electronic products such as computers and cell phones, mainly including platform applications, website applications, and cell phone applications. These health management applications with health data recording, online health consultation, and health condition assessment make health management no longer restricted by time and place and give greater play to the value of medical resources, achieving the effect of reducing disease rates and improving the quality of life [18]. As governments continue to launch exploratory national health management construction documents and plans, major Internet companies have responded to the development of the times and have invested in research in the health management field.

So far, many advanced and professional health management applications have been launched abroad. These include the platform applications Health Kit and Google Fit, the website application Health Vault, and a large number of cell

phone applications, among which cell phone applications have the advantages of portability and easy operation and are most easily received and used by people [19].

Medisafe reminds users to take their medication on time through push messages and prompts them to record the amount of medication they take when they take it [20]. If the system does not detect that the patient has taken the medication, patient's friends and family are notified and can take action against the patient. In addition, Sweetech, an Israeli company, has developed a disease risk diagnosis and prevention software, which can predict user's chances of getting sick and make users who are already sick but do not know if they are aware of their condition. In addition to Western countries, health management systems are also growing rapidly in Asian countries such as Korea and Japan. Samsung's S Health has become one of the most accepted personal health management platforms in Korea today. S Health mainly includes environmental detection, exercise management, and health diary, through which users can record their daily health activities and customize their health and exercise plans to accomplish their goals under the guidance of the system [21]. Physical health data such as heart rate can also be measured through the sensors built into the Samsung Galaxy and Gear. In 2017, Japanese healthcare company OmronHealthcare invested in Kardia Pro, a heart disease prevention management software from Alivecor Alive Cor, which tracks users' ECG, weight, blood pressure, daily behavior, etc., uses artificial intelligence technology for data analysis and comparison to identify potential risk factors that are not detected by doctors, and organizes the results in user's personal health profile to help doctors diagnose the user in case of sudden illness.

### 3. Big Data-Based Physical Health Management Model

*3.1. A Closed-Loop Health Management Model Based on Big Data.* Closed-loop management is a complete management system, first proposed by Robert S. Kaplan, founder of the Balance Card, and initially applied to improve the efficiency of company management and operations. With the development of health management field, the traditional simple health management model cannot meet the needs of people's health management, and closed-loop management in health management has been widely used. Closed-loop health management mainly refers to the process of decision-making, implementation, feedback, reimplementation, and feedback in health management, so that personal physical quality can be continuously improved in the cycle of accumulation and promote the improvement or maintenance of personal health level. The closed-loop health management system mainly targets people with subhealth or subdisease, through the steps of health data collection, health assessment, plan making, guidance intervention, and user feedback, cultivating users' good living habits, improving their self-health management awareness and self-healing power, and achieving the effect of controlling the occurrence or deterioration of diseases at the root. Professor Pan Xiaojue from the Shanghai Jiaotong University pointed out that

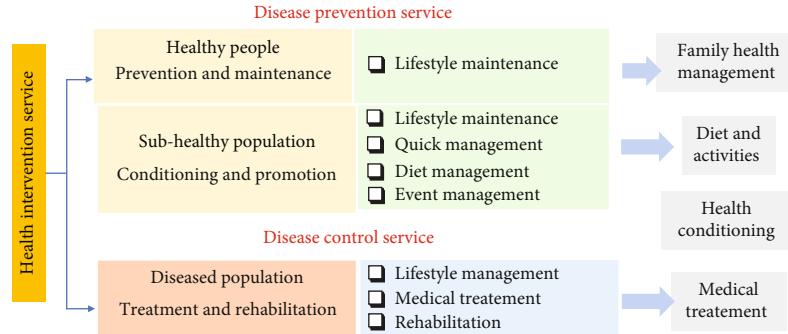


FIGURE 2: Closed-loop health management model.

health management mainly starts from three closed loops, the first one is the closed loop of testing means, body's physiological indicators are constantly changing, and in the process of health management, physiological indicators are constantly tested to find out which physiological indicators have problems, so as to solve problems. The second is the closed-loop intervention path; closed-loop health management system is mainly used by people with chronic diseases, not only from the consideration of improving the disease but also through the system to continuously intervene to guide the lifestyle and gradually improve the root cause of the disease. The third is the closed loop of repairing raw materials. The repairing raw materials of chronic diseases and sub-healthy people are mainly their own health management consciousness and mentality, and a good health management consciousness is a strong immunity and self-healing power. Zhang Juan et al. from the Tongji Medical College conducted a study on closed-loop health management for diabetic people, proposing a dynamic closed-loop diabetes health management model of prehospital prevention-in-hospital education-posthospital rehabilitation through online diagnosis and treatment, medication guidance, and access to or sharing of health data, etc.

The researchers explored the information-based closed-loop health management model for subhealthy people under the Internet+ model, in which medical and nursing personnel realize long-term intelligent closed-loop health management for subhealthy people in and out of the hospital through health data monitoring [22], trend analysis, guidance intervention, and online tracking and management. The PDCA management cycle, proposed by Dr. Hart, an American quality management expert, divides management into four stages: plan, do, check, and act. The PDCA management cycle is a very effective management model for all kinds of management, which can develop people's own management awareness and make users develop the instinct of automatic management through the steps of planning guidance and result checking. The PDCA management cycle is applied to personal health management system, which is the basic management model of personal health management system based on multidimensional data model proposed in this paper, to achieve closed-loop personal health management through the cycle of data collection, data collation, health assessment, guidance intervention, and imple-

mentation of programs. The personal health management model proposed in this paper is shown in Figure 2.

**3.2. Big Data Collection.** Standardized big data collection and recording methods are the basis for health management of chronic diseases and subhealthy people and are the basic requirements for data collation and analysis, health assessment, personalized health intervention, and management effect evaluation. There are various channels and types of health data collection, mainly from professional clinical data of medical institutions, health data integrated by regional health service platforms, conventional health data collected by portable wearable devices, and users' self-quantified health data. The clinical data of medical institutions refers to professional medical health data, which includes medical data generated from various medical clinical consultations and medical examinations, and these health data are the most valuable data in users' health management. Regional health service platform fully brings together personal health data from many medical institutions of users, which at this stage mainly refers to medical checkup data. Portable wearable devices mainly collect users' ECG, blood pressure, blood oxygen, sleep, pulse, body temperature, and other physical signs data. User self-quantified health data includes exercise information, diet, smoking, and drinking. Self-quantified data is also an important basis for systematic health assessment and disease judgment, which helps to provide personalized health guidance and intervention to users. These health data are generally divided into three formats: structured, semistructured, and unstructured. In this paper, by collecting the above channels and types of health data, we analyze and mine the health data through distributed data processing technology and perform trend analysis, health prediction, and risk warning on user's health status in the system.

Based on health data sources, this paper is designed to obtain users' health information through the following three paths: PC-based website system, mobile-based APP platform, and smart wearable devices.

(1) PC-based website system

According to the classification, format, and other data attributes of health data, we design various health data

record forms based on the PC-based health management website. The collection method based on the PC-based website is based on user-initiated entry, and the basic principles of health data collection refer to the relevant regulations in the Basic Architecture and Data Standards for Health Records and Basic Architecture and Data Standards for Electronic Medical Records issued by the Ministry of Health of China. Among them, the basic personal information collection form refers to “HRA00.01 Basic Data Set Standard for Personal Information,” the medical physical examination data entry form refers to “HRC00.04 Basic Data Set Standard for Adult Health Examination,” the hypertension collection record module refers to “HRB04.01 Basic Data Set Standard for Hypertension Case Management,” and the diabetes collection record module refers to “HRB04.02.” The other health data collection record modules are also designed according to the data format and attributes in the standard.

### (2) Mobile-based APP platform

The main form of the personal health management system described in this paper is based on the mobile APP platform, and this health management system has developed APPs based on both Android and iOS in order to facilitate users to record and manage their health at any time. In addition to the PC-based website system, the mobile APP is designed with interfaces for unstructured data voice, image, audio, and video collection. The database between the PC-based website system and the mobile-based APP is synchronized through the server database, which facilitates user records and unified data management.

### (3) Smart wearable-based devices

Smart wearable devices are the main tools for daily monitoring of body data for patients with chronic diseases and are the main source of routine health data. Examples include real-time wearable health devices such as dynamic blood glucose monitors and ECG monitors, as well as nonreal-time monitoring devices such as pedometers and blood pressure monitors. The detected health information is transmitted and stored in the database of this system through the Bluetooth function that smart health devices have.

The design of the database model is the focus of this paper. The traditional personal health management system does not conduct in-depth research on the attributes of health data such as diseases and body organs. This paper proposes a personal health management system based on multidimensional data model, which adopts a snowflake multidimensional data model and mines the attributes of diseases, body organs, and clinical departments of health data according to relevant clinical diagnosis guidelines and health data standards issued by the Ministry of Health of China, so that users can manage health data, intelligent health assessment, and disease risk prediction from multiple dimensions such as diseases and body organs, fully reflecting the value of health. The value of health data is fully reflected.

The traditional personal health management system only collects a large amount of personal health data and makes a

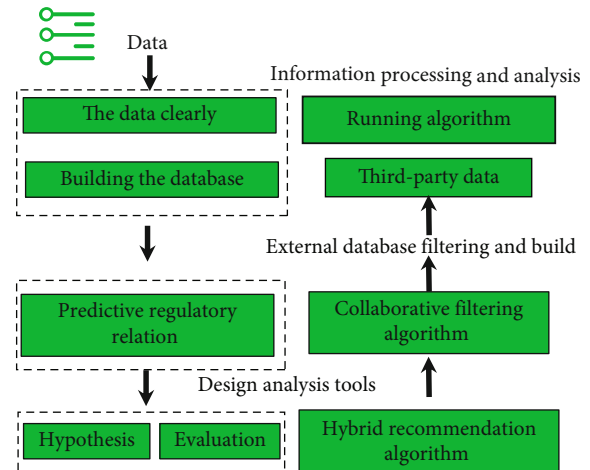


FIGURE 3: Multidimensional warehouse application for big data.

simple list, without taking into account the multiple attributes and dimensions of health data, so that the potential value of health data in health assessment and disease prediction cannot be fully utilized. Usually, a health data has multiple attributes and can be used for diagnosis of multiple body parts and disease types, for example, common blood pressure data; a large amount of blood pressure data can be used not only for judging the level of blood pressure but also for prediction assessment of cardiovascular system diseases combined with heart rate, daily symptoms and living habits, and for judging cerebrovascular health status combined with intraocular pressure and brain symptoms. In this paper, we design a multidimensional database model and set various attributes such as diseases and organs for each health data element, so that users can flexibly choose the use of health data and decide the role of data, so that health data can play a greater value in health assessment and disease prediction of personal health management system. In the personal health management system based on multidimensional data model, the relationship between each health data type needs to be standardized and designed according to the health assessment analogy, and clear and accurate data relationships and data attributes help to improve the decision-making accuracy of personal health management health assessment and disease prediction, etc. This paper designs a data warehouse with multidimensional data model, as shown in Figure 3, which mainly studies and solves the problem of obtaining information from the database. It features a topic-oriented, integrated, stable, and time-varying data collection, which is used to analyze data and discover new values of data, mainly for analytical reports and decision support in management, and meets the needs of user health data analysis and evaluation in health management systems.

## 4. Artificial Intelligence-Based Health Fuzzy Integrated Assessment Model

4.1. *Health Assessment Index Weights*. Indicator weights represent the importance of the indicator in health assessment,

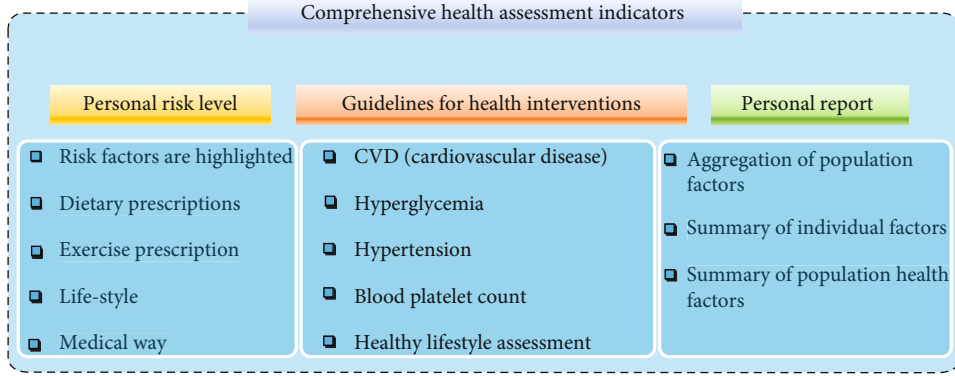


FIGURE 4: Hierarchy of health assessment indicators.

and reasonable allocation of each assessment indicator weight is the key to ensure the accuracy and scientificity of health assessment, and this paper selects hierarchical analysis method to obtain the assessment indicator weights. Hierarchical analysis was proposed by Thomas L. Saaty, an American operations researcher, in the 1970s. It is a combination of qualitative and quantitative decision analysis method, which can convert semiquantitative and semiquantitative problems into quantitative calculations. The health evaluation system built in this paper has both qualitative and quantitative evaluation factors, so using this method can hierarchize the complex medical health problem system and calculate the importance or weight value of each factor quantitatively by comparing the importance between related factors to establish a preliminary quantitative evaluation system. The main steps of the hierarchical analysis method to obtain the weights of evaluation indicators are as follows: firstly, the complex decision problem is stratified according to the general objective, criterion layer, and indicator layer, and then, the judgment matrix of each layer is determined, and then, the priority weights of each indicator in each layer to the elements of the upper layer are obtained by solving the eigenvectors of the judgment matrix and performing the consistency test.

### (1) Layering

According to the hierarchical analysis method, the hierarchical structure of the comprehensive health assessment index system was determined. As shown in Figure 4, the hierarchical structure of health assessment indexes is divided into target layer A; criterion layer B, with two categories; primary index layer C, with 9 items; and secondary index layer D, with 6 items.

### (2) Constructing the judgment matrix

After completing the evaluation model of evaluation indicators, the indicators within each level are analyzed and compared with each other, and the judgment matrix of the corresponding level is constructed. Judgment matrix construction method: select a layer of indicators in the hier-

archy, and experts compare the importance of the indicators in that layer relative to the indicators in the upper layer two by two and record the comparison results in the form of a matrix, which can constitute the judgment matrix between two levels. In this paper, 10 experts are selected to participate in the comparison of the importance of each index, taking into account the relevance of experts' research field to the topic, the time of experts' practice, and the age distribution of experts. In this paper, 1-9 scale is used to quantitatively scale the importance between two indicators, and 1, 3, 5, 7, and 9 are used to indicate the importance of one indicator over another, and the larger the number means the more important it is; when an indicator is less important than another indicator, the reciprocal of 1, 3, 5, 7, and 9 is used to indicate the less important degree; when the hierarchical degree needs higher precision, the real numbers between 1 and 9 can be interpolated to improve the precision.

The judgment number  $a_{ij}$  represents the assignment of the importance of index  $A_i$  over  $A_j$ , and the judgment matrix of each layer of order  $nm^*$  is formed by  $a_{ij}$ . 10 experts determine the judgment numbers of two comparisons of each factor between the indexes of a layer according to the scale of 1-9 and average the judgment numbers given by the experts to obtain the judgment matrix of each layer as shown in

$$Q = \begin{bmatrix} q_{11} & q_{12} & \cdots & q_{1i} \\ \cdots & & & \\ q_{i1} & q_{i2} & \cdots & q_{ii} \end{bmatrix}, \quad (1)$$

$$Q_{|A-B|} = \begin{bmatrix} 22 \\ 11/2 \end{bmatrix}, \quad (2)$$

$$Q_{|C-B|} = \begin{bmatrix} 2 & 2 & 1 & 1/5 \\ 1 & 1/2 & 3 & 1/4 \\ 2 & 3 & 4/5 & 6 \\ 1/3 & 3 & 2 & 1 \end{bmatrix}. \quad (3)$$

(3) Hierarchical single ordering and its consistency detection

After constructing the judgment matrix, the hierarchical single ranking and consistency verification are performed according to the judgment matrix, and the maximum eigenvalue  $\lambda_{\max}$  of each layer judgment matrix is found out by MATLAB, and the normalized eigenvector of the matrix is found out, and each component  $\lambda_i$  of the eigenvector is the important ranking weight of this layer index relative to the upper layer index. In order to reflect the rationality and objectivity of the indicator weights, this paper conducts consistency verification for each judgment matrix constructed to test whether the matrix of each layer meets the requirements. The steps of the consistency test are as follows:

$$c_I = \frac{(\lambda_{\max} - n)}{(\sqrt{n + m})}. \quad (4)$$

The deviation consistency index indicates the degree of deviation of the judgment matrix from the consistency matrix, where  $\lambda_{\max}$  denotes the maximum eigenvalue of the judgment matrix and  $n$  denotes the dimensionality of the judgment matrix:

$$Cr = \frac{c_I}{R} = \frac{(\lambda_{\max} - n)}{R(\sqrt{n + m})}. \quad (5)$$

Since the variability among the layers increases with the superposition of the layers, it is also necessary to conduct a consistency test on the total ranking, and the weight values that pass the total ranking consistency test can be determined to be adopted. The total ranking consistency test method: let the matrix of the relevant evaluation indexes  $M_i (i = 1, 2, \dots, n)$  in  $M + 1$  layer and  $M$  layer pass the test of the matrix passed the single-sort consistency test, the single-sort deviation consistency index of the matrix is  $CI_i$ , the average random consistency index is  $RI_i$ , and the weight value of  $M_i$  is  $i$ . Then, the relative consistency index of the total sort in  $M$  layer can be calculated by the formula:

$$R = \sum_{i=1}^n \frac{Cr_i}{Rw_i}. \quad (6)$$

Bringing in the values to calculate the relative consistency index of the primary indicator layer relative to the criterion layer, the results are shown in

$$R = \frac{0.1}{0.9} + \frac{0.2}{0.5} + \frac{0.3}{0.6} = 1.02. \quad (7)$$

Fuzzy comprehensive evaluation refers to the overall evaluation of things that are affected by multiple evaluation factors and have fuzzy nature by using the idea of fuzzy mathematics. It has the characteristics of clear purpose, strong system, and clear structure and can reasonably quantify the fuzzy things that are difficult to evaluate objectively in real life. In this paper, we choose the fuzzy comprehensive

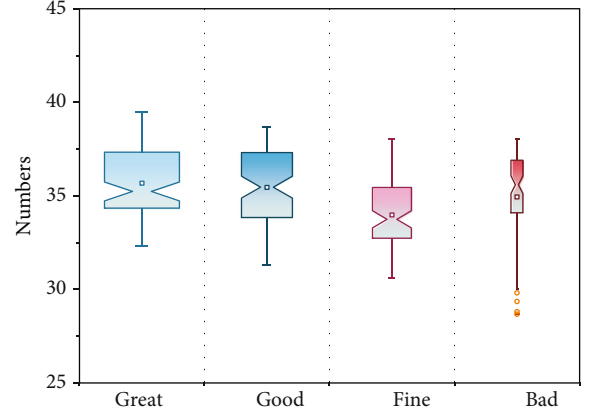


FIGURE 5: Diagram of blood glucose subordination and subordination function cut-off points.

evaluation method to evaluate the human health status. After obtaining the evaluation index weights by the above hierarchical analysis method, we use the affiliation function to construct the fuzzy matrix, the index weight set and the fuzzy evaluation matrix constitute the fuzzy subcomprehensive evaluation set, and finally, according to the principle of maximum affiliation, we get the health evaluation results. In this paper, we choose to construct the trapezoidal distribution affiliation function according to the data characteristics of health assessment indexes. The following is an example of blood glucose index to analyze the affiliation function, as Figure 5 shows the diagram of blood glucose affiliation and affiliation function dividing point.

The affiliation function with the rank of superior: here, the  $Q$  means the judgment matrix, and the  $q_1$  and  $q_2$  means the judgment constant, which can be obtained from Figure 5:

$$Q = \begin{cases} 1, & x \leq q_1, \\ \frac{(q_2 - q_1)}{2}, & q_1 < x < q_2, \\ 0, & x \leq q_2. \end{cases} \quad (8)$$

The affiliation function for a grade of good:

$$Q = \begin{cases} 2, & x \leq q'_1, \\ \frac{(q'_2 - q'_1)}{2}, & q'_1 < x < q'_2, \\ 1, & x \leq q'_2. \end{cases} \quad (9)$$

The rank is the affiliation function in:

$$Q = \begin{cases} 3, & x \leq q'_{1'}, \\ \frac{(q'_{2'} - q'_{1'})}{2}, & q'_{1'} < x < q'_{2'}, \\ 2, & x \leq q'_{2'}. \end{cases} \quad (10)$$



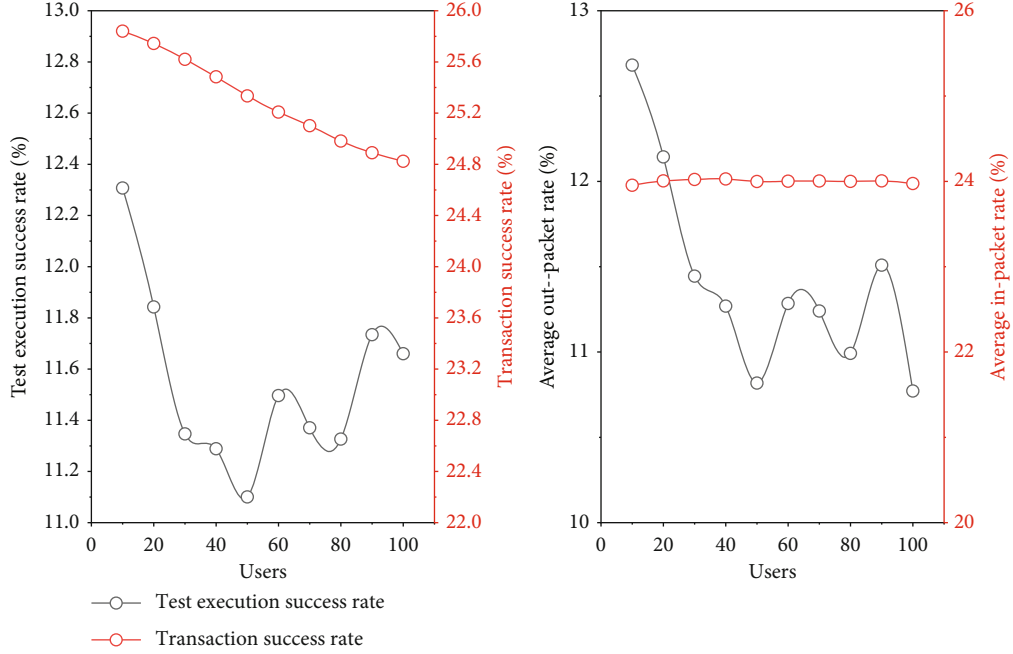


FIGURE 6: Performance test results under different load conditions.

The rank is a poor affiliation function:

$$Q = \begin{cases} 4, & x \leq q'''_1, \\ \frac{(q'''_2 - q'''_1)}{2}, & q'''_1 < x < q'''_2, \\ 3, & x \leq q'''_2. \end{cases} \quad (11)$$

## 5. System Testing




**5.1. System Performance Testing.** The system was tested for software performance using Pressure Test Master, and the software performance such as the success rate of system time request, average transaction response time, packet sending rate, and packet receiving rate were tested, respectively. Meanwhile, the compatibility of the system was also tested, and the compatibility and adaptability of APP were tested using different brands of smartphones, respectively.

Figure 6 shows the summary of user system performance test results when the load concurrent users are 10, 20, 30, and 50, and the test results in the table show that the system transaction response success decreases slightly as the concurrent users increase, but the overall response time is above 90%; the average transaction response time is less than 1 second, which indicates that the system response speed is stable; the packet sending rate and packet receiving rate show a positive relationship, and the packet receiving rate continues to increase as the hit rate increases. As the hit rate increases, the packet receipt rate continues to increase, indicating that the server can respond to user event requests in a timely manner. The system responds well to the concurrent access of multiple users and meets the performance requirements of the user software. In this paper, the comprehensive

health assessment model is studied and applied, and the fuzzy comprehensive assessment algorithm based on AHP is used to design the comprehensive health assessment model of this system, and the assessment model is tested using examples.

**5.2. Health Data Collation.** Health data recording is the C (check) step in the PDCA closed-loop health management model, and it is also a common function module for users to use the system on a daily basis. Users of the system can record health data through two channels: the website on the PC side and the APP health record on the mobile terminal. In addition to the text input on the website, the mobile terminal APP data recording method also has picture upload, Bluetooth import, voice recognition, and other recording methods. Figure 7 shows an example of the classification of the health data recording module in APP terminal, which is the same as the classification basis and database table in website terminal. According to user's demand for one-stop health management, a closed-loop health management model based on the PDCA cycle management model is adopted to design and implement the system functions from four aspects: health data collection and recording, health assessment, health plan, and tracking and execution.

The health plan is the P (plan) step in the PDCA closed-loop health management model. The user makes a health improvement plan based on the health assessment results obtained in the above three ways. The plan form includes the contents that need to be recorded for a long time, such as the plan name, target data elements, and target values, and the contents that can be tracked and reminded by the system, such as the number of exercise steps and the name and time of medication. The system judges the

			
1900/1/1	0.36	0.36	
1900/1/2	0.36	0.37	
1900/1/3	0.35	0.36	
1900/1/4	0.35	0.36	
1900/1/5	0.34	0.34	
1900/1/6	0.36	0.35	
1900/1/7	0.36	0.36	
1900/1/8	0.35	0.36	
1900/1/9	0.36	0.35	
1900/1/10	0.35	0.36	
1900/1/11	0.35	0.36	
1900/1/12	0.34	0.36	
1900/1/13	0.36	0.35	
1900/1/14	0.36	0.34	
1900/1/15	0.35	0.35	
1900/1/16	0.37	0.35	
1900/1/17	0.36	0.34	
1900/1/18	0.35	0.36	
1900/1/19	0.37	0.35	
1900/1/20	0.35	0.36	
1900/1/21	0.37	0.34	
1900/1/22	0.37	0.35	
1900/1/23	0.35	0.35	

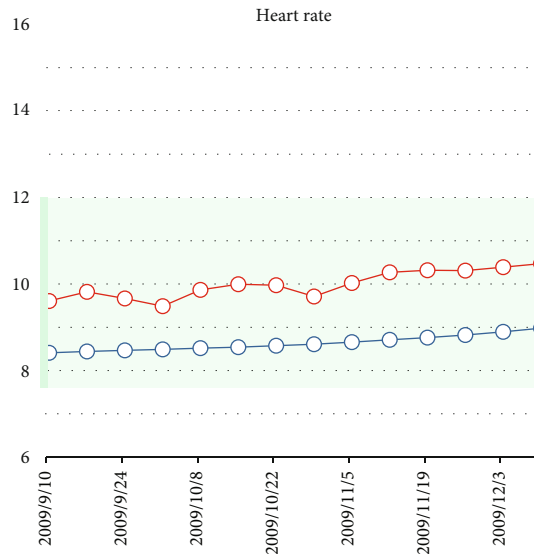


FIGURE 7: Health data summary presentation.

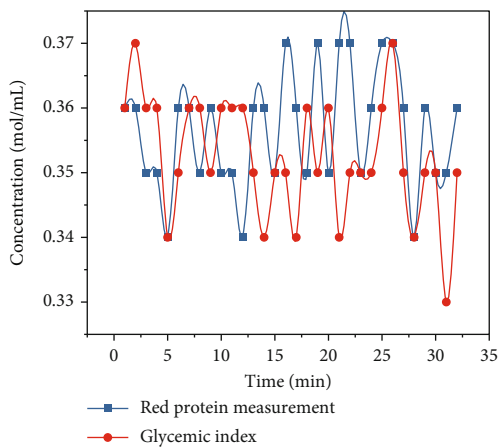


FIGURE 8: Glycemic index change graph.

reasonableness of user’s health plan according to user’s specific health condition and guides the user to make a suitable health plan. Figure 8 shows the effect of developing a diabetes control management plan. User’s health plan is generated into a plan book, and the system will detect the items in the plan book that need to be tracked and reminded and then realize the next step of tracking and execution.

## 6. Conclusion

With the increasing incidence of chronic diseases in China, people’s awareness of chronic disease management has been strengthened, and the emergence of personal health management systems is the general trend. At present, most of the health management systems in China have a single function and cannot meet the needs of long-term health management of users. In this paper, we summarize the shortcomings of

existing health management systems and the needs of chronic disease population health management, and design and implement a personal health management system based on multidimensional data model. The system can meet users’ various health management needs, create personalized health management guidance program for users, and achieve the purpose of improving users’ health in all aspects. The average transaction response time is less than 1 second, which indicates that the system response speed is stable; the packet sending rate and packet receiving rate show a positive relationship, and the packet receiving rate continues to increase as the hit rate increases. According to user’s demand for one-stop health management, a closed-loop health management model based on the PDCA cycle management model is adopted to design and implement the system functions from four aspects: health data collection and recording, health assessment, health plan, and tracking and execution. In this paper, the comprehensive health assessment model is studied and applied, and the fuzzy comprehensive assessment algorithm based on AHP is used to design the comprehensive health assessment model of this system, and the assessment model is tested using examples. This paper uses software development and database development techniques to design and implement this system, initially realizing the functions of login and registration, health data collection, health assessment, health planning, and tracking execution. The functions and performance of the software system were also tested, and the test results showed that this system basically meets user’s needs for a health management system.

In the future, we would explore the information-based closed-loop health management model for subhealthy people under the Internet+ model, in which medical and nursing personnel realize long-term intelligent closed-loop

health management for subhealthy people in and out of hospital through health data monitoring, trend analysis, guidance intervention, and online tracking and management.

### Data Availability

The data used to support the findings of this study are included within the article.

### Conflicts of Interest

The author declares that there is no conflict of interest.

### References

- [1] J. Wang, Z. Xie, Y. Li et al., "Relationship between health status and physical fitness of college students from south China: an empirical study by data mining approach," *IEEE Access*, vol. 8, pp. 67466–67473, 2020.
- [2] F. Zhao, S. Sun, J. Xiong, and G. Zheng, "The effect of Baduanjin exercise on health-related physical fitness of college students: study protocol for a randomized controlled trial," *Trials*, vol. 20, no. 1, pp. 569–569, 2019.
- [3] X. Lu, C. Yang, Y. Zhang et al., "Test method for health-related physical fitness of college students in mobile internet environment," *Mathematical Biosciences and Engineering*, vol. 16, no. 4, pp. 2189–2201, 2019.
- [4] S. Mal-Ryun, "The comparative analysis of skeletal muscle mass, basal metabolic rate abdominal obesity and health related physical fitness of female college students - targeting students who majored in dance and general," *Official Journal of the Koeran Society of Dance Science*, vol. 37, no. 1, pp. 29–38, 2020.
- [5] T. Singh, "Impact of yogic asanas on the health-related physical fitness of students of agricultural biotechnology college," *International Journal of Physical Education, Sports and Health*, vol. 4, no. 1, pp. 2250–2253, 2019.
- [6] C. B. Yang and T. H. Tsao, "Cardiorespiratory fitness, health-related physical fitness and academic performance in college students," *Advances in Physical Education*, vol. 10, no. 1, pp. 42–53, 2020.
- [7] M. Sawant, "Efficacy of yogic practices for the promotion of academic achievement, mental health and health related physical fitness of college level female students," *International Journal of Yoga, Physiotherapy and Physical Education*, vol. 4, no. 4, pp. 14–17, 2019.
- [8] T. S. Eshete, W. Mekonen, and H. Derseh, "Comparative assessment of vegetarian and non-vegetarian diets with physical fitness on body composition and lipid profiles among students at School of Medicine, College of Health Sciences, Addis Ababa University," *Asian Journal of Medicine and Health*, vol. 13, no. 4, pp. 1–15, 2019.
- [9] P. Tang, Y. Wang, and N. Shen, "Prediction of college students' physical fitness based on K-means clustering and SVR," *Computer Systems: Science & Engineering*, vol. 35, no. 4, pp. 237–246, 2020.
- [10] P. Bo, "Analysis and intervention on the influencing factors of college students' physical fitness," *Revista Brasileira De Medicina Do Esporte*, vol. 27, pp. 11–13, 2021.
- [11] J. L. Wang, "The association between physical fitness and physical activity among Chinese college students," *Journal of American College Health*, vol. 67, no. 6, pp. 602–609, 2019.
- [12] T. Pan, "An improved a priori algorithm for association mining between physical fitness indices of college students," *International Journal of Emerging Technologies in Learning (IJET)*, vol. 16, no. 9, pp. 235–246, 2021.
- [13] C. Qian, "Changes in body composition, blood pressure, bone density and physical fitness of female college students by wearing wearable devices for 24 weeks," *The Korean Society of Sports Science*, vol. 29, no. 1, pp. 683–694, 2020.
- [14] S. I. Mushtaq and A. Kumar, "Comparative study of selected physical fitness components between urban and rural college level students," *International Journal of Physical Education, Sports and Health*, vol. 4, no. 1, pp. 2403–2406, 2019.
- [15] T. M. Barnett, A. McFarland, J. W. Miller, V. Lowe, and S. S. Hatcher, "Physical and mental health experiences among African American college students," *Social Work in Public Health*, vol. 34, no. 2, pp. 145–157, 2019.
- [16] D. A. Kinney, L. A. Nabors, A. L. Merianos, and R. A. Vidourek, "College students' use and perceptions of wearable fitness trackers," *American Journal of Health Education*, vol. 50, no. 5, pp. 298–307, 2019.
- [17] J. Mora-Gonzalez, I. J. Pérez-López, and M. Delgado-Fernández, "The 'Sin TIME' gamification project: using a mobile app to improve cardiorespiratory fitness levels of college students," *Games for Health Journal*, vol. 9, no. 1, pp. 37–44, 2020.
- [18] M. Wang, Y. Guo, Y. Zhang et al., "Promoting healthy lifestyle in Chinese college students: evaluation of a social media-based intervention applying the RE-AIM framework," *European Journal of Clinical Nutrition*, vol. 75, no. 2, pp. 335–344, 2021.
- [19] P. Prieto-González and J. Sedlacek, "Physical education: practice, cessation and resumption in untrained Saudi college students," in *Series Physical Education and Sport*, pp. 681–692, Facta Universitatis, 2021.
- [20] N. A. Mir, "Comparative study of health related physical fitness of college level students of volleyball and football players," *International Journal of Physical Education, Sports and Health*, vol. 7, no. 6, pp. 222–224, 2020.
- [21] J. DiMatteo, C. Radnitz, K. L. Loeb, and J. Ni, "The application of optimal defaults to physical education courses in college students: a simulation study," *Journal of Teaching in Physical Education*, vol. 38, no. 4, pp. 393–397, 2019.
- [22] L. Yu, S. Tao, W. Gao, and L. Yu, "Self-monitoring method for improving health-related quality of life: data acquisition, monitoring, and analysis of vital signs and diet," *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 1, pp. 24–31, 2021.

## Research Article

# A Novel Stock Index Intelligent Prediction Algorithm Based on Attention-Guided Deep Neural Network

Yangzi Zhao <sup>1,2</sup>

<sup>1</sup>Sichuan Industrial and Commercial College, Meishan 620010, China

<sup>2</sup>Business School, University of Dundee, Nethergate, Dundee, Scotland, UK

Correspondence should be addressed to Yangzi Zhao; [ukzhaoyangzi2018@163.com](mailto:ukzhaoyangzi2018@163.com)

Received 5 August 2021; Revised 25 August 2021; Accepted 27 August 2021; Published 21 September 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Yangzi Zhao. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The stock market is affected by economic market, policy, and other factors, and its internal change law is extremely complex. With the rapid development of the stock market and the expansion of the scale of investors, the stock market has produced a large number of transaction data, which makes it more difficult to obtain valuable information. Because deep neural network is good at dealing with the prediction problems with large amount of data and complex nonlinear mapping relationship, this paper proposes an attention-guided deep neural network stock prediction algorithm. This paper synthesizes the daily stock social media text emotion index and stock technology index as the data source and applies them to the long-term and short-term memory neural network (LSTM) model to predict the stock market. The stock emotion index is extracted by constructing a social text classification emotion model of bidirectional long-term and short-term memory neural network (Bi-LSTM) based on attention mechanism and glove word vector representation algorithm. In addition, a dimensionality reduction model based on decision tree (DT) and principal component analysis (PCA) is constructed to reduce the dimensionality of stock technical indicators and extract the main data information. Furthermore, this paper proposes a model based on nasNet for pattern recognition. The recognition results can be used to automatically identify short-term K-line patterns, predict reliable trading signals, and help investors customize short-term high-efficiency investment strategies. The experimental results show that the prediction accuracy of the proposed algorithm can reach 98.6%, which has high application value.

## 1. Introduction

The stock market is an important part of a country's economy, which seriously affects the formulation of individual and national investment strategies. Program trading is the development trend of the future stock exchange market [1]. The formulation of automation strategy, as an important part of program trading, directly affects the long-term and short-term investment income [2]. The data-driven stock market forecast provides a more reliable buying and selling signal for the automatic trading strategy, which can maximize the user's investment income. The stock market is a complex and nonlinear environment, which is affected by many variable factors, mainly including five aspects: (1) economic variables, (2) company-specific variables, (3) factory-specific variables, (4) political variables, and (5) investor psychological variables. How to successfully predict the change trend of

the stock market and capture the behavior pattern of the stock market in such a complex stock market environment is regarded as the most meaningful and challenging task [3]. As an indispensable part of individual and national economy, the stock market has been a hot topic for a long time. In the field of data analysis, compared with earlier studies, researchers realize that the stock market is a whole composed of a large number of stocks, and there is a high correlation between stock indexes; at the same time, the latest development of sensor networks and communication technology makes it possible to collect massive stock data, so how to effectively process massive stock data, successfully predict the change trend of stock market, and capture the behavior mode of stock market has become the focus of research [4].

However, the accuracy of stock forecasting is limited by many factors. Because stock data is a random walk financial time series, literature [5] demonstrates that the difficulty of

financial time series prediction lies in its high noise. If we use a statistical model to predict financial time series, we must preprocess the data and input it into the model, which will destroy the integrity and authenticity of the data. Secondly, stock data is nonlinear and nonstationary. Literature [6] believes that the nonlinearity and nonstationary of stock data lead to the limited application of traditional multiple regression and linear regression models. Only higher-level models can accurately describe such nonlinear financial time series. Finally, the stock market is affected by many complex and uncertain factors, such as long-term macro policies, short-term market expectations, and some emergencies or international events. Therefore, the deep neural network method based on attention is used to predict the stock in the paper.

## 2. Related Work

*2.1. Stock Forecasting Based on Regression Method.* Scholars at home and abroad have tried many methods to predict the stock market. Literature [7] used the vector autoregressive (VaR) model, error correction model (ECM), and Kalman filter model (KFM) to predict the UK stock market in 1996. Literature [8] uses the Bayesian vector autoregressive model (BVaR) to predict the portfolio return of some large German companies, but the prediction effect is poor. Based on the stock data of the New York Stock Exchange and Nigeria stock exchange, literature [9] attempts to use the ARIMA model to predict stock prices. The results show that the model has short-term prediction potential [10]. However, because the stock data is a nonlinear and nonstationary financial time series, and there are many and complex factors affecting the stock price, the traditional statistical methods and measurement models need to preprocess the input data, and the amount of data cannot be too much, resulting in the unsatisfactory prediction effect. Literature [11] uses the support vector machine (SVM) with good generalization ability and fast computing ability to predict stock market prices. Literature [12] applies the gray system model to the prediction of China's stock market and obtains reasonable and reliable results. Literature [13] uses the model integrating genetic algorithm (GA) with high convergence and artificial neural network (ANN) to predict the stock market. In addition to the basic machine learning algorithm, researchers also try to use the new algorithm framework based on machine learning to predict the stock market. Among them, literature [14] uses K-neighbor neural network (Knn-Bp) to predict the stock market. Literature [15] uses a fuzzy model with the characteristics of avoiding empirical subjectivity and selecting objectivity to predict the opening price, closing price, maximum price, and minimum price of the stock market every day. However, early studies only considered some simple stock influencing factors, resulting in relatively low accuracy of stock prediction [16].

*2.2. Stock Forecasting Based on Neural Network.* Literature [17] studies have found that the accuracy of the neural network model [18–20] in predicting nonlinear time series data

is much higher than the ARIMA model. Literature [21] compares the prediction effects of the Bayesian estimation and neural network model with different standards. The results show that the prediction effect of the neural network model is better. Literature [22] established the AR model, RBF, and GRNN neural network models to predict the opening price, closing price, highest price, and lowest price of Shanghai stock index and compared with the actual price and analysed the error, demonstrating the effectiveness of the three models, but the AR model is relatively unstable, RBF and GRNN network training speed is very fast, and GRNN shows a better effect [23].

Literature [24] proposed a model integrating optimized bacterial chemotaxis (IBCO) and back propagation (BP) artificial neural network algorithm to effectively predict various stock indexes. However, the early artificial neural network has a large number of parameters [25], which is prone to problems such as fitting and gradient dispersion, so the prediction accuracy is not high. With the introduction of dropout and other structures in neural network, deep neural network has become a hot spot in predicting stock market [26]. Among them, the depth learning models with good performance include convolutional neural network (CNN), classical cyclic neural network (RNN), and cyclic convolutional neural network (RCNN) [27]. Among them, LSTM, a kind of cyclic neural network, can integrate long information and short information well and solve the problems of gradient dispersion and gradient explosion. LSTM is widely used in time series prediction [28]. Literature [29] developed a powerful adaptive online gradient learning algorithm based on LSTM to predict time series with outliers. Literature [30] combines LSTM depth network and basic statistical algorithm to predict multistep time series.

## 3. Stock Index Prediction Algorithm Based on Attention-Guided Deep Neural Network

*3.1. Structure Design for Stock Index Prediction Algorithm.* Many studies have pointed out that investors' emotional indicators and stock technical indicators are positively correlated with the changes of the stock market, so using the emotional indicators extracted from social media and traditional technical indicators to predict the stock market has become a research hot spot [31]. In terms of technology, the excellent performance of in-depth learning in natural language and time series tasks makes it possible to successfully predict the stock market. In this chapter, a system model is constructed to predict the stock price and change trend by using the historical emotional indicators and technical indicators with a period of days, so as to provide reliable prediction results for long-term investors and help investors formulate high-efficiency investment strategies. The system model is shown in Figure 1.

As can be seen from Figure 1, this paper constructs a Bi-LSTM [32] text classification model based on attention and glove. The trained model is used to predict the emotional classification of stock-related social media texts in real time and extract emotional indicators from the prediction results. This chapter also constructs a dimensionality reduction

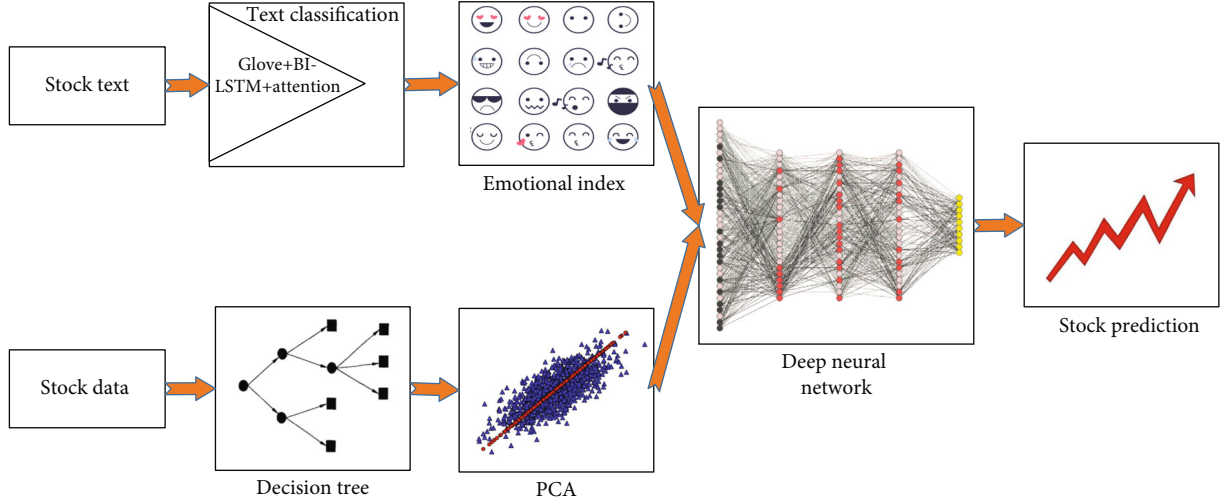


FIGURE 1: Stock forecasting algorithm structure figure.

model [33] based on decision tree and principal component analysis to reduce the dimensionality of stock technical indicators. Finally, the index is applied to the LSTM model to predict the price and change trend of the stock market. Use the historical emotion index and technical index of the stock with a period of days to predict the stock price and change trend, so as to provide reliable prediction results for long-term investors.

**3.2. Structure Design of Social Media Text Emotion Classification Model.** Text emotion classification model based on two-way long-term and short-term memory neural network based on focus mechanism and glove. The model structure is shown in Figure 2. The classification model is composed of one globe layer, two Bi-LSTM layers, and one attention layer.

For a stock-related social media text sequence  $x$ , because the LSTM structure is introduced into the classification model, it is necessary to truncate the  $X$  sequence into a fixed length  $m$ , where  $m$  is the concept of “memory length” in the LSTM structure. As can be seen from Figure 2, the text sequence  $x$ , as the input of the glove layer, is output as a vector matrix representing words. The experimental results show that the word vector classification result of 200 dimensions is the most accurate. Therefore, the vector dimension of a word is selected as 200 dimensions in this paper. As the first layer of the model, the glove algorithm combines the advantages of latent semantic analysis algorithm (LSA) and continuous word bag algorithm (CBOW). It has faster training speed and better scalability for large-scale corpus algorithm. The loss function of glove algorithm can be defined as

$$J = \sum_{ik} f(X_{ik})(w_i^T w_k + b_i + b_k - \log X_{ik}), \quad (1)$$

where  $X_{ik}$  represents the number of occurrences of word  $k$  in the context of word  $i$ , which is a weight function used to

weigh the influence between two words, where  $w_k$  and  $w_i$  represent word  $k$  and word  $i$ , respectively.  $b_i$  and  $b_k$  are the offset term.

The trained stock-related social media emotion classification model and the predicted emotion categories will provide the stock emotion index; it is

$$\begin{cases} \text{Pos}R_t = \frac{M_t^{\text{pos}}}{(M_t^{\text{pos}} + M_t^{\text{neg}})}, \\ \text{Neg}R_t = \frac{M_t^{\text{neg}}}{(M_t^{\text{pos}} + M_t^{\text{neg}})}, \end{cases} \quad (2)$$

where  $M_t^{\text{pos}}$  and  $M_t^{\text{neg}}$  mean the number of texts in which investors have positive and negative attitudes towards a stock.

**3.3. Method Design of Bi-LSTM.** The output of the glove layer will be used as the input of the Bi-LSTM layer. As the core structure of Bi-LSTM, LSTM is a kind of cyclic neural network structure. Classical recurrent neural network in “memory” super.

As can be seen from Figure 3, LSTM is composed of a group of connection blocks called “memory blocks,” which can be regarded as a differentiated version of digital computer memory. LSTM is mainly composed of forgetting gate, input gate, output gate, and storage unit.

$$\begin{aligned} f_t &= \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \\ i_t &= \sigma(W_C \cdot [h_{t-1}, x_t] + b_C), \\ \tilde{C}_t &= \tanh(W_C \cdot [h_{t-1}, x_t] + b_C), \\ \tilde{C}_t &= \tanh(W_C \cdot [h_{t-1}, x_t] + b_C), \\ o_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \\ h(t) &= o_t * \tanh(C_t), \end{aligned} \quad (3)$$

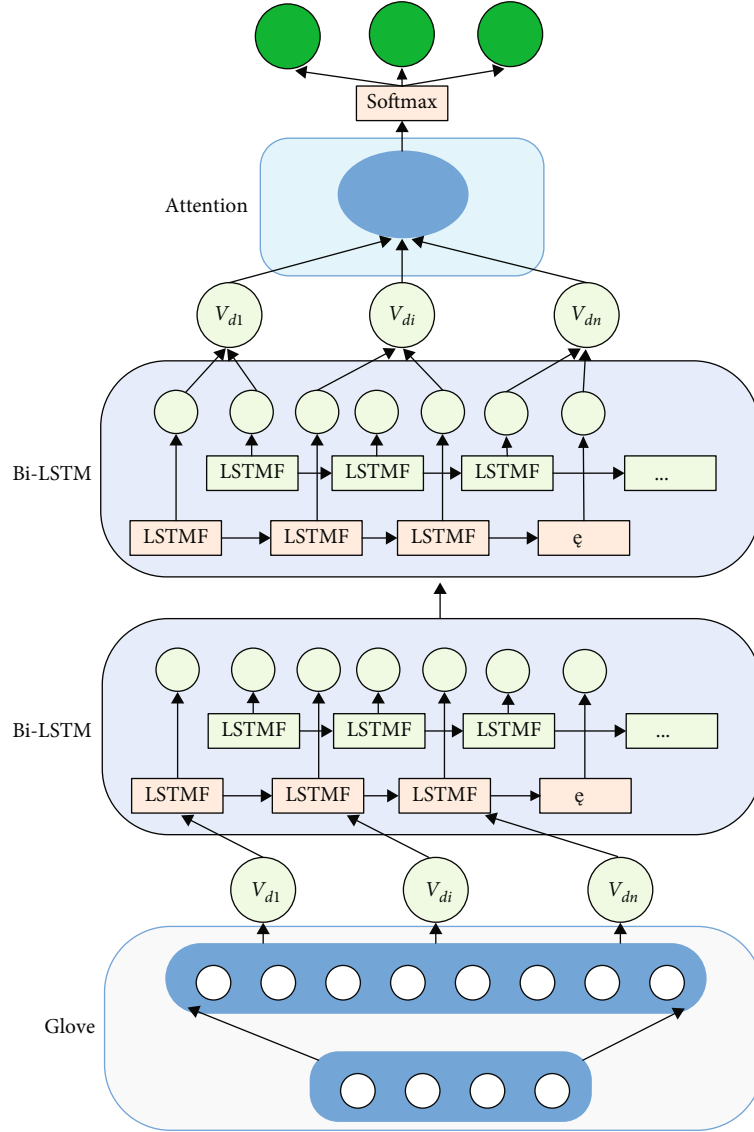


FIGURE 2: Social media text emotion classification model based on Bi-LSTM attention.

where  $f_t$  and  $i_t$  are the intermediate functions of forgetting gate and input gate, respectively, and  $o_t$  and  $c_t$  are the output functions of the state gate and the output gate, respectively.  $h_t$  is the output function of LSTM neurons. In addition,  $W_f$ ,  $W_i$ ,  $W_o$ , and  $W_c$  represent the weight parameters of the network, respectively.

**3.4. Attention Layer Design for Stock Prediction.** As the fourth layer of the classification model, the attention mechanism is introduced in this chapter. Natural language processing operates by “reading” a complete sentence and compressing the sentence information into a fixed-length vector. It is conceivable that information loss, incomplete transformation, and other problems will occur in a sentence compression near a low dimensional vector composed of hundreds of words, and the attention mechanism solves these problems to a certain extent. It allows the machine to traverse the whole sentence information and then produce reasonable results according to the current word and the

whole sentence. The attention layer is designed as in Figure 4.

$$\alpha_{ts} = \frac{\exp(\text{score}(h_t, \tilde{h}_s))}{\sum \exp(\text{score}(h_t, \tilde{h}_s))}. \quad (4)$$

The attention mechanism first calculates  $\alpha_{ts}$ .  $h_t$  and  $\tilde{h}_s$  represent the target state and source state, respectively. In this classification model, the source state SH is the output of the bidirectional LSTM layer in step  $t$ , and  $s$  is all the states generated by the bidirectional LSTM. The score function is designed to balance the influence of weights.

$$\alpha_t = \tanh(W_c[c_t, h_t]). \quad (5)$$

After the data passes through the attention layer and passes through a simple softmax layer, the emotion types

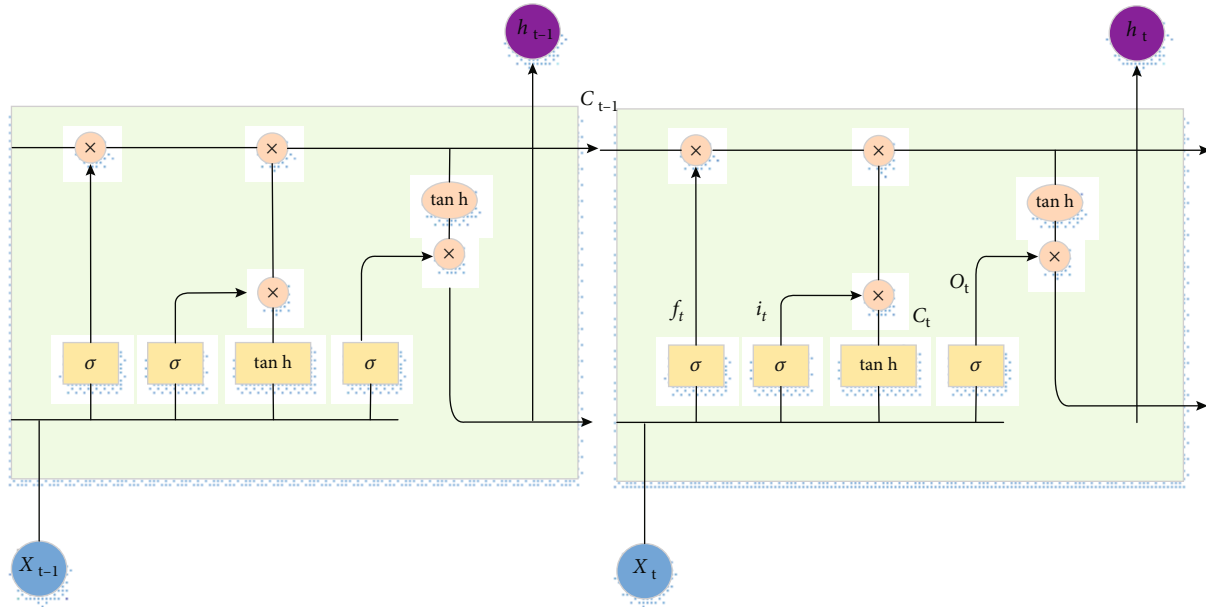


FIGURE 3: The structure of LSTM neural network.

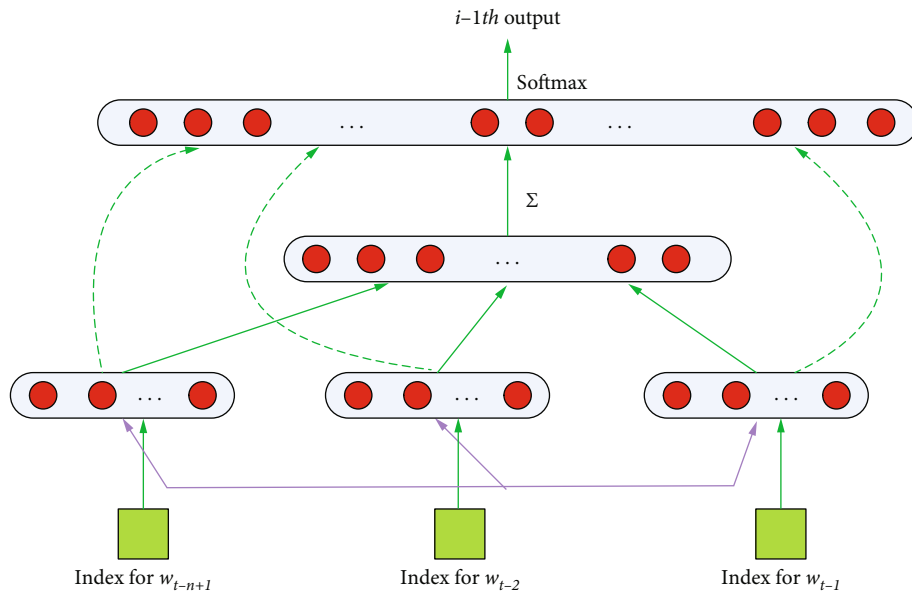


FIGURE 4: The index structure of attention layer.

of social media texts can be predicted. So far, a complete emotion classification model of stock-related social media texts has been constructed. Through the training of cost function and back propagation algorithm on the training set, a trained emotion classification model can classify social media texts in real time.

#### 4. Short-Term Stock Prediction Model Based on nasNet

4.1. Structure Design of nasNet for Stock Prediction. After years of research, stock researchers have pointed out that the short-term minute K-line pattern chart can provide use-

ful trading signals and help investors formulate efficient investment strategies. However, the investment strategies based on K-line mode now need to be captured by manual observation or hard coding. The former consumes manpower, and the latter has limited range setting and cannot capture the K-line mode diagram flexibly. Therefore, a method that can automatically and comprehensively capture the K-line mode diagram is needed, and an automatic and high-benefit strategy is provided according to the captured results.

The image processing model based on convolutional neural network can get good results in high-speed calculation on large-scale training set, but the amount of calculation



is very large. In order to train a good model on a small training set and get good results when migrating to a large training set, the author of the nasNet model proposes to build an RNN controller to automatically generate two structural units: normal cell and reduction cell. These two units will be used as basic units to build a complete model together with basic convolutional neural network units. Figure 5 shows the structure of the nasNet model for stock prediction. As can be seen from the figure, the network structure of nasNet is only several layers more than that on small dataset cifar10, but it is composed of normal unit, restore unit, and volume layer unit as basic units, which fully shows that nasNet structure can be simply migrated from small dataset to large dataset; the implementation of this structure significantly reduces the amount of calculation on large datasets and reduces the calculation cost.

**4.2. Structure of Reduction Cell and Normal Cell.** Normal cell and reduction cell are the most important core cells in nesNet. Therefore, the two cell structures need to be described in detail. The structure of normal cell and reduction cell is shown in Figure 6.

In the nasNet structure, so it can be seen in Figure 6 that there are five addition operations in both the normal unit and the restore unit. In addition, the structure does not directly connect unused hidden layers in series. On the contrary, all hidden layers created in the convolution unit are sent to the next layer even if they are currently used.

## 5. Experiment and Result Analysis

**5.1. Experimental Environment and Dataset Source.** In order to facilitate the statistics of the experimental results, I selected 10 U.S. stocks in different industries as the experimental objects, including Amazon (AMZN), Apple (AAPL), Facebook (FB), Google (Google), Microsoft (MSFT), Netflix (NFLX), qq, S&P (spy), Twitter (TWTR), and Tesla (tsla). The social media platform used to extract emotional indicators in this experiment is Twitter. In this experiment, a total of 10 stocks were crawled and 7.3 million tweets were used as the social media data source to extract emotional indicators in 1461 trading days; in order to obtain reliable emotion indicators, this paper constructs a social media text emotion classifier to improve the accuracy of emotion prediction. Therefore, it is necessary to mark some tweets as training sets and test sets. A total of 12670 tweets are marked in this experiment, which are divided into three emotion types. At the same time, the reliability of the extracted emotion indicators will be further guaranteed in the future; you need to ensure that each stock has at least 500 tweets a day.

This experiment uses the first 6650 data as the training set and the remaining 121 data as the test set. The data indicators include seven indicators: opening price, closing price, highest price, lowest price, daily trading volume, rise and fall range, and turnover rate, which correspond to the characteristic dimension at each time in the model. This paper uses all seven indicators of the day to predict tomorrow's closing price.

Tensorflow is selected as the experimental platform, and the experimental environment is a computer equipped with GTX 1080ti graphics card and 32G memory. The initial learning rate of the model is 0.0007, with a total of 2000 rounds of learning (one round represents that all samples in the training set participate in training once). Firstly, the data are preprocessed, the mean value under each feature dimension is subtracted and then divided by the standard deviation to obtain the standardized data, and then, the data are input into the model for training.

**5.2. Verification of Emotion Index Classification Results.** In this chapter, 12670 stocks are selected for emotion classification and labeling. In order to maintain the balance of data and the adjustability of model parameters, the number of texts in each category is equivalent, including 4753 positive texts, 4703 negative texts, and 4215 neutral texts, respectively; in addition, the experiment shows that when the truncation length of each text is 20, the classification result is more accurate. Therefore, in this part of the experiment, the truncation length of the text is set to 20. In the experiment, 70% was selected as the training set and 30% as the test set; at the same time, in order to evaluate the performance of the social media text emotion classification model, accuracy, recall, precision, and  $F$  value are selected as evaluation measures. In order to verify the effectiveness of the social media text emotion classification model, Bi-LSTM, CNN+Bi-LSTM, and Glove+Bi-LSTM models are constructed as control. The experimental results are shown in Table 1.

As can be seen from Table 1, compared with other methods, the social media text emotion classification model constructed in this chapter performs best in the test set (accuracy 0.7659, recall 0.7282,  $F$  value 0.7663, and accuracy 0.75). Compared with Bi-LSTM and Glove+Bi-LSTM models, the social media text emotion classification model constructed in this paper fully illustrates the role of glove algorithm and attention mechanism in natural language classification task. The accuracy of the CNN+Bi-LSTM model is 7% lower than that of the constructed classification, mainly because the social media text is relatively short and colloquial, while CNN structure is good at dealing with long text structure.

In order to intuitively display the classification results of three emotional texts (positive text, negative text, and neutral text), I define neutral text as 0, positive text as positive, and negative text as negative. In this way, the classification results of emotional text based on attention-Glove-Bi-LSTM are shown in Figure 7. Figures 7(a) and 7(b), respectively, show the classification results, Figure 7(c) shows the classified heat energy, and Figure 7(d) shows the clustering probability. As can be seen from the figure, the above algorithm can cluster three types of emotional texts, indicating the effectiveness of the clustering algorithm. Therefore, this paper constructs a social media text emotion classification instrument, which has a high accuracy of emotion prediction. As shown in the figure, the positive, negative, and neutral classification and clustering have obvious boundaries,

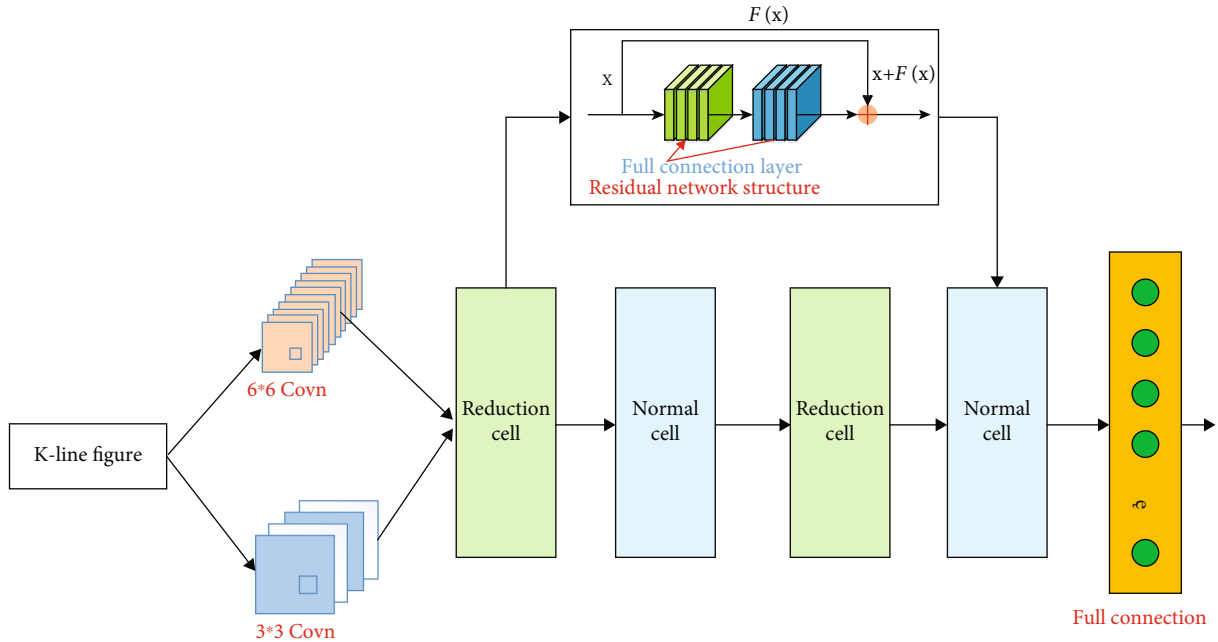


FIGURE 5: The structure design of nasNet.

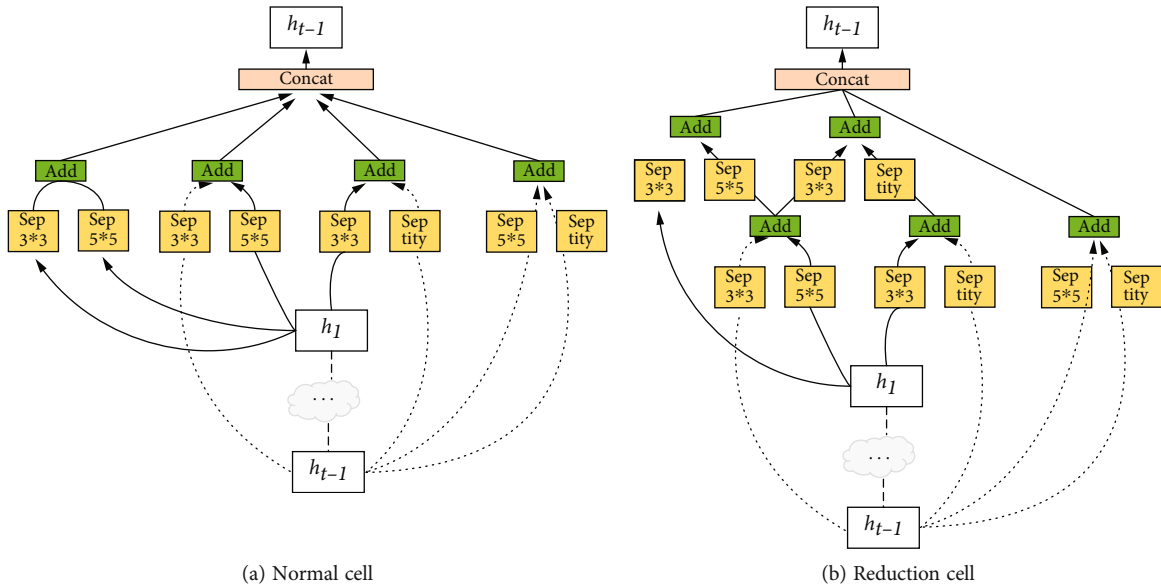


FIGURE 6: The structure of normal cell and reduction cell.

TABLE 1: Social media text prediction.

Methods	Accuracy	Recall	F value	Precision
Bi-LSTM	0.726	0.684	0.657	0.597
CNN+Bi-LSTM	0.734	0.647	0.672	0.675
Glove+Bi-LSTM	0.752	0.675	0.708	0.724
Attention+Glove+Bi-LSTM	0.768	0.782	0.743	0.76

which shows that the emotional text classification algorithm proposed in this paper has high classification accuracy.

5.3. Relationship between CNN Model Performance and nasNet Structure. This section explores the impact of these important structural parameters on the performance of the model by adjusting the structural parameters of the nasNet model. Firstly, the convolution kernel in Figure 5 is replaced by the square convolution kernel often used in the image task, and 2 are used, respectively,  $2 \times 2$ ,  $3 \times 3$ ,  $4 \times 4$ , and 5

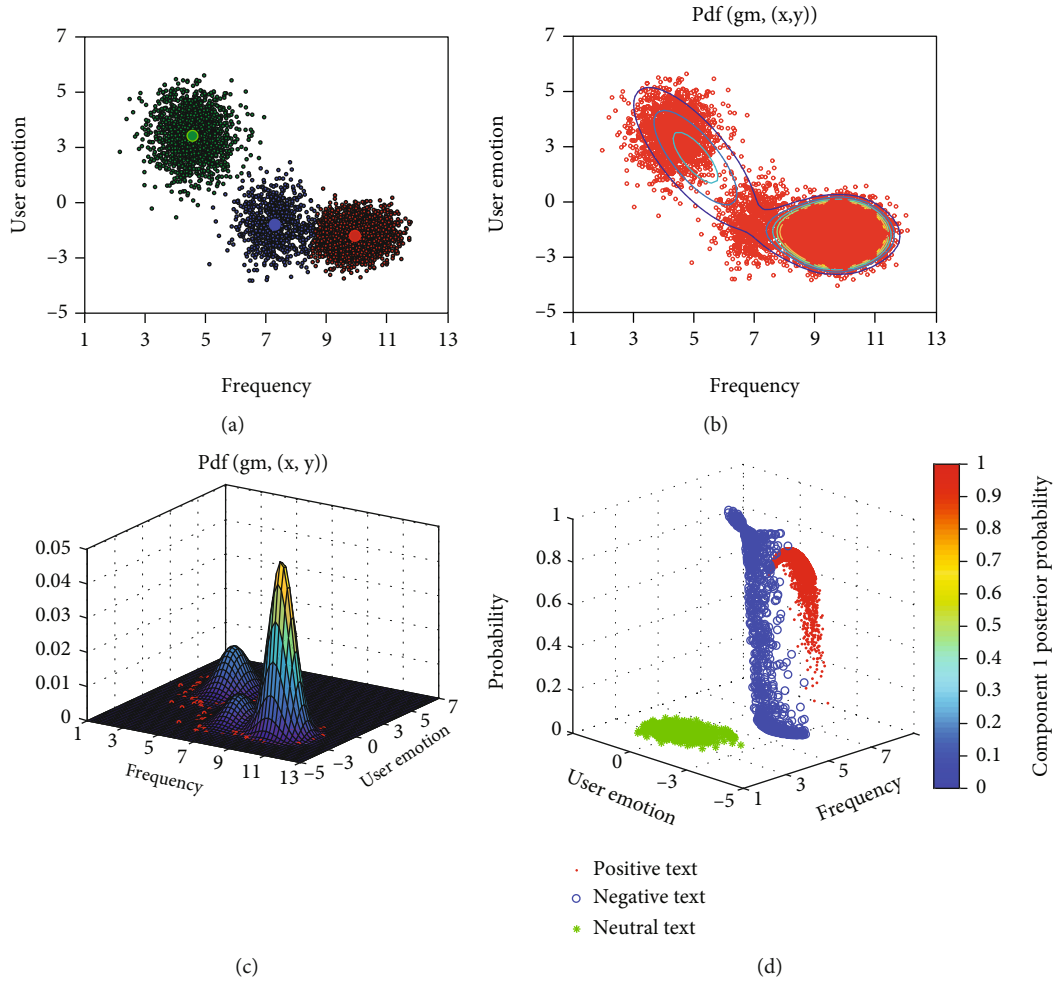


FIGURE 7: The clustering results based on attention-Glove-Bi-LSTM: (a) classification results; (b) classification results with contour line; (c) thermal energy diagram; (d) probability figure of user emotion.

$\times 5$ . The prediction curve of the model is shown in Figure 8. As you can see,  $2 \times$  the convolution kernel of 2 performs the worst, and the average error is as high as 1.6357%, which is much higher than the other three convolution kernels. This is because it is more difficult to extract the global information of convolution check data with too small size, especially the stock data with significant global correlation. The convolution kernel of 5 shows the best effect, which may be because the larger convolution kernel can more richly extract the correlation information between the feature dimensions. However, the performance of all the above models cannot be compared with the optimized CNN model, which proves the effectiveness of the convolution kernel optimization method for data form.

As can be seen in Figure 8, with the increase of convolution blocks, the model performance shows a downward trend. There are two main reasons: on the one hand, the stock data structure is relatively simple, and the complexity is far less than that of image data. Therefore, using a relatively simple network structure can approach the upper limit of the model, and a more complex network will only lead to

serious overfitting. On the other hand, CNN is not suitable for processing time series tasks. The results show that the convolution kernel of  $3 \times 3$  has strong prediction ability and prediction accuracy.

As shown in Figure 8, the short-term K-line pattern recognition model based on nasNet is used to automatically identify four common short-term K-line patterns. The experimental results show that the model can well identify four common short-term K-line patterns, and the accuracy is 98.6%. This result can not only free investors from the heavy observation of stock market changes and help investors customize automatic high-efficiency investment strategies but also provide a reliable trading signal for the construction of automatic trading strategies in the future.

**5.4. Accuracy Verification of Stock Forecast.** Stock market price forecasting is very important for investors to make strategies. Therefore, in this part of the experiment, I mainly study the prediction of emotional indicators and technical indicators on the closing price of the stock market, and the prediction of nesNet on the time series of stock data. In this

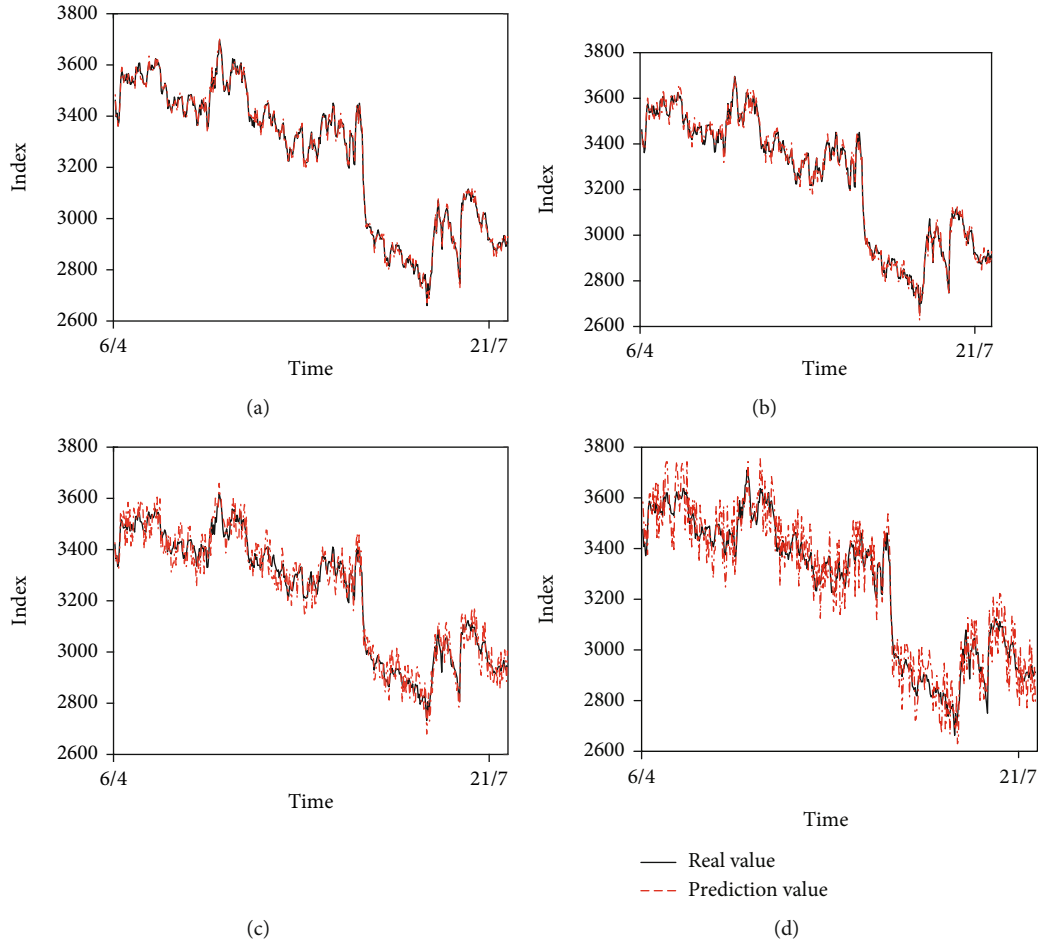


FIGURE 8: Model prediction performance under different convolution kernel structures: (a)  $2 \times 2$  convolution kernel; (b)  $3 \times 3$  convolution kernel; (c)  $4 \times 4$  convolution kernel; (d)  $5 \times 5$  convolution kernel.

TABLE 2: Performance comparison with other methods.

Methods	RMSE	MAE	MAPE (%)	Performance			
				$R$	$R^2$	MSE	Ranking
LSTM	0.41	0.26	1.40	0.51	0.24	0.17	2
Ridge	0.51	0.35	1.90	0.60	0.35	0.26	3
kneighbors	1.67	1.20	6.28	1.10	1.20	2.19	4
Decision tree	3.11	2.68	15.75	1.64	2.68	9.67	5
SVM	3.44	3.31	18.76	1.82	3.31	11.83	6
Improved nesNet	0.35	0.20	1.13	0.39	0.20	0.13	1

part of the experiment, a total of 6 emotional indicators and 82 technical indicators are introduced, a total of 88 indicator characteristics are used to predict the changes of the stock market. In order to improve the prediction accuracy, the heat map method is used to select indicators with a correlation coefficient of no less than 0.9 with the closing price to predict the closing price of the stock, and a total of 22 indicator characteristics are selected. These 22 indicators will be used as inputs to the nesNet structure to predict the closing price. In order to evaluate the experimental results of this

part, the statistical method root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), mean square error (MSE), correlation coefficient ( $R$ ), and nonlinear regression multiple correlation coefficient ( $R^2$ ) are selected as the evaluation indexes. The formulas of these indexes are shown in Table 2, where  $y$  represents the real value of the closing price,  $\tilde{y}$  represents the predicted value of the closing price,  $\bar{y}$  represents the average value of the real price in the test set, and  $n$  represents the number of test sets. The model based on nasNet is used to identify

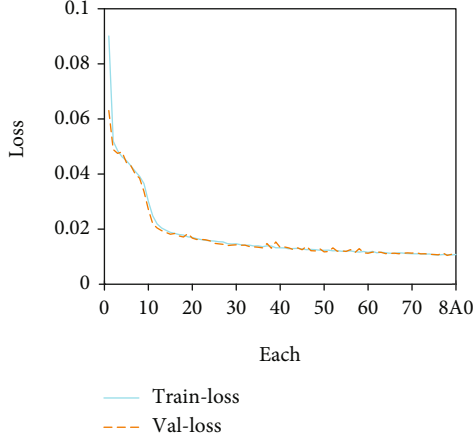


FIGURE 9: Training loss figure under nasNet structure.

the above K-line pattern. The loss value of the specific training process on the test set is shown in Figure 9.

$$\begin{aligned}
 \text{RMSE} &= \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y} - y)^2}, \\
 \text{MAE} &= \frac{1}{n} \sum_{i=1}^n |y - \hat{y}|, \\
 \text{MAPE} &= \frac{1}{n} \sum_{i=1}^n \left| \frac{y - \hat{y}}{y} \right|, \\
 \text{RMSE} &= \frac{1}{n} \sum_{i=1}^n (\hat{y} - y)^2, \\
 R &= \frac{n \sum y \hat{y} - \sum y \sum \hat{y}}{\sqrt{n \sum y^2 - (\sum y)^2} \sqrt{n \sum \hat{y}^2 - n \sum \hat{y}^2}}.
 \end{aligned} \tag{6}$$

It can be clearly seen from Figure 9 that using the model parameters trained in advance to train the nasNet-based image recognition model can quickly achieve the convergence effect, and the loss values of the training set and the test set are very low. In order to evaluate the K-line pattern recognition ability of the nasNet model on small-scale datasets, the accuracy is selected as the evaluation standard. At the same time, in order to verify the role of the model, the experiment selects the latest recognition algorithm in the image field in recent years, in order to verify the advantages of nasNet in small dataset image recognition.

In order to verify the effect of the improved nesNet model on the prediction of stock closing price with time series characteristics, LSTM model, ridge regression, neighbors, decision tree, and support vector machine (SVM) algorithm are introduced as the comparison group. In addition, the improved nesNet model needs to fine tune many parameters. In order to ensure the whole experimental process and comparability, the experiment selects the parameters with the best prediction results according to the experimental results and fixes these settable parameters as constant values. The output is set to 128 units, and the disconnection degree

is set to 0.2. For the models of the control group and the experimental group, the 1461 day trading day data from January 1, 2014, to December 31, 2017, are used as the original data, and 70% are set as the training group and 30% as the control group. The experimental results are shown in Table 2.

In the stock closing price prediction experiment, the model proposed in this paper performs best (RMSE: 0.35, MAE: 0.20, MAPE: 1.13,  $R$ : 0.39,  $R^2$ : 0.20, and MSE: 0.13). As can be seen, the accuracy of the improved nesNet model in predicting the closing price is much higher than that of the control group algorithm, which fully illustrates the excellent performance of the improved nesNet model in the time series prediction task. In future research, the improved nesNet model can also be used to predict more stock market trends.

## 6. Conclusion

In the field of stock forecasting, the research on data-driven stock forecasting is of great significance to the stock market and the development of automation in the future. Taking social media text and stock technical indicators as data sources, this paper constructs a two-way long and short memory model based on attention mechanism and glove word representation and a dimensionality reduction model based on principal component analysis and decision tree fusion to extract stock emotion indicators from social media text and reduce the dimensionality of stock technical indicators. The extracted stock sentiment index/dimensionality reduction technical index and K-line chart are used as the input of the improved nesNet model to predict the long-term stock market price and stock change trend, respectively. The experimental results show that the prediction method of stock market price and stock change trend proposed in this paper can effectively improve the prediction accuracy. In addition, compared with other methods, it further verifies the superiority of the improved nesNet model considering emotional text. However, the stock market is a complex and nonlinear environment. There is still much room for development in terms of prediction accuracy and strategy formulation, so further research is needed in the future.

## Data Availability

The data used to support the findings of this study are included within the article.

## Conflicts of Interest

The author does not have any possible conflicts of interest.

## References

- [1] X. Zhang, Y. Hu, K. Xie, S. Wang, E. W. T. Ngai, and M. Liu, "A causal feature selection algorithm for stock prediction modeling," *Neurocomputing*, vol. 142, pp. 48–59, 2014.

- [2] H. Chung and K.-s. Shin, "Genetic algorithm-optimized long short-term memory network for stock market prediction," *Sustainability-Basel*, vol. 10, no. 10, p. 3765, 2018.
- [3] Y. Chen and Y. Hao, "A feature weighted support vector machine and K-nearest neighbor algorithm for stock market indices prediction," *Expert Systems with Applications*, vol. 80, pp. 340–355, 2017.
- [4] F. Wang, Y. Zhang, Q. Rao, K. Li, and H. Zhang, "Exploring mutual information-based sentimental analysis with kernel-based extreme learning machine for stock prediction," *Soft Computing*, vol. 21, no. 12, pp. 3193–3205, 2017.
- [5] R. H. Gálvez and A. Gravano, "Assessing the usefulness of online message board mining in automatic stock prediction systems," *Journal of Computational Science*, vol. 19, pp. 43–56, 2017.
- [6] L. Zhang, F. Wang, B. Xu, W. Chi, Q. Wang, and T. Sun, "Prediction of stock prices based on LM-BP neural network and the estimation of overfitting point by RDCI," *Neural Computing and Applications*, vol. 30, no. 5, pp. 1425–1444, 2018.
- [7] G. H. Kim and S. H. Kim, "Variable selection for artificial neural networks with applications for stock price prediction," *Applied Artificial Intelligence*, vol. 33, no. 1, pp. 54–67, 2019.
- [8] L. Tang, P. H. Pan, and Y. Y. Yao, "EPAK: a computational intelligence model for 2-level prediction of stock indices," *International Journal Of Computers Communications & Control*, vol. 13, no. 2, pp. 268–279, 2018.
- [9] Z. Dai and H. Zhou, "Prediction of stock returns: sum-of-the-parts method and economic constraint method," *Sustainability-Basel*, vol. 12, no. 2, p. 541, 2020.
- [10] M. Nabipour, P. Nayyeri, H. Jabani, A. Mosavi, and E. Salwana, "Deep learning for stock market prediction," *Entropy*, vol. 22, no. 8, article e22080840, p. 840, 2020.
- [11] R. Bhowmik, R. J. Berry, M. F. Durstock, and B. J. Leever, "Prediction of the wetting behavior of active and hole-transport layers for printed flexible electronic devices using molecular dynamics simulations," *ACS applied materials & interfaces*, vol. 9, no. 22, pp. 19269–19277, 2017.
- [12] J.-S. Chou and T.-K. Nguyen, "Forward forecast of stock price using sliding-window metaheuristic-optimized machine-learning regression," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 7, pp. 3132–3142, 2018.
- [13] Y. Wang and Y. Guo, "Forecasting method of stock market volatility in time series data based on mixed model of ARIMA and XGBoost," *China Communications*, vol. 17, no. 3, pp. 205–221, 2020.
- [14] C. B. Musgrave III, K. Kim, N. R. Singstock, A. M. Salazar, J. W. Stansbury, and C. B. Musgrave, "Computational and experimental evaluation of peroxide oxidants for amine-peroxide redox polymerization," *Macromolecules*, vol. 53, no. 22, pp. 9736–9746, 2020.
- [15] P. P. S. Kohli, S. Zargar, S. Arora, and P. Gupta, "Stock prediction using machine learning algorithms," *Applications of Artificial Intelligence Techniques in Engineering*, vol. 1, pp. 405–414, 2019.
- [16] F. Yang, Z. Chen, J. Li, and L. Tang, "A novel hybrid stock selection method with stock prediction," *Applied Soft Computing*, vol. 80, pp. 820–831, 2019.
- [17] A. Thakkar and K. Chaudhari, "Fusion in stock market prediction: a decade survey on the necessity, recent developments, and potential future directions," *Information Fusion*, vol. 65, pp. 95–107, 2021.
- [18] H. D. Huynh, L. M. Dang, and D. Duong, "A new model for stock price movements prediction using deep neural network," in *Proceedings of the Eighth International Symposium on Information and Communication Technology*, pp. 57–62, New York, NY, USA, 2017, December.
- [19] I. T. Joo and S. H. Choi, "Stock prediction model based on bidirectional LSTM recurrent neural network," *The Journal of Korea Institute of Information, Electronics, and Communication Technology*, vol. 11, no. 2, pp. 204–208, 2018.
- [20] Y. Tong, L. Yu, S. Li, J. Liu, H. Qin, and W. Li, "Polynomial fitting algorithm based on neural network," *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 1, pp. 32–39, 2021.
- [21] W. Chen, H. Zhang, M. K. Mehlatat, and L. Jia, "Mean-variance portfolio optimization using machine learning-based stock price prediction," *Applied Soft Computing*, vol. 100, article 106943, 2021.
- [22] W. Lu, J. Li, J. Wang, and L. Qin, "A CNN-BiLSTM-AM method for stock price prediction," *Neural Computing and Applications*, vol. 33, no. 10, pp. 4741–4753, 2021.
- [23] H. Wang, S. Lu, and J. Zhao, "Aggregating multiple types of complex data in stock market prediction: a model-independent framework," *Knowledge-Based Systems*, vol. 164, pp. 193–204, 2019.
- [24] N. Jing, Z. Wu, and H. Wang, "A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction," *Expert Systems with Applications*, vol. 178, no. 3, article 115019, 2021.
- [25] M. Miao, W. Cai, and X. Li, "Parameter estimation of gamma-gamma fading with generalized pointing errors in FSO systems," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 1301878, 14 pages, 2021.
- [26] M. Göçken, M. Özçalıcı, A. Boru, and A. T. Dosdoğru, "Stock price prediction using hybrid soft computing models incorporating parameter tuning and input variable selection," *Neural Computing and Applications*, vol. 31, no. 2, pp. 577–592, 2019.
- [27] F. Giannetti, G. Chirici, T. Gobakken, E. Næsset, D. Travaglini, and S. Puliti, "A new approach with DTM-independent metrics for forest growing stock prediction using UAV photogrammetric data," *Remote Sensing of Environment*, vol. 213, pp. 195–205, 2018.
- [28] H. Kong, Y. Fang, L. Fan, H. Wang, X. Zhang, and J. Hu, "A novel torque distribution strategy based on deep recurrent neural network for parallel hybrid electric vehicle," *IEEE Access*, vol. 7, pp. 65174–65185, 2019.
- [29] X. Pang, Y. Zhou, P. Wang, W. Lin, and V. Chang, "An innovative neural network approach for stock market prediction," *The Journal of Supercomputing*, vol. 76, no. 3, pp. 2098–2118, 2020.
- [30] S. K. Chandar, "Grey wolf optimization-Elman neural network model for stock price prediction," *Soft Computing*, vol. 25, no. 1, pp. 649–658, 2021.
- [31] X. Li, P. Wu, and W. Wang, "Incorporating stock prices and news sentiments for stock market prediction: a case of Hong Kong," *Information Processing & Management*, vol. 57, no. 5, article 102212, 2020.

- [32] S. Mootha, S. Sridhar, R. Seetharaman, and S. Chitrakala, "Stock price prediction using bi-directional LSTM based sequence to sequence modeling and multitask learning," in *2020 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, pp. 0078–0086, New York, NY, USA, 2020, October.
- [33] Z. Wang, P. Zhang, W. Sun, and D. Li, "Application of data dimension reduction method in high-dimensional data based on single-cell 3D genomic contact data," *ASP Transactions on Computers*, vol. 1, no. 2, pp. 1–6, 2021.

## Research Article

# Neural Network Topology Construction and Classroom Interaction Benchmark Graph Based on Big Data Analysis

Congcong Luan <sup>1,2</sup> and Peng Shang<sup>3</sup>

<sup>1</sup>School of Marxism, Shandong University, Jinan 250000, China

<sup>2</sup>Zibo Normal College, Zibo 255100, China

<sup>3</sup>Boshan Middle School, Zibo 255100, China

Correspondence should be addressed to Congcong Luan; [lcc820125@163.com](mailto:lcc820125@163.com)

Received 27 July 2021; Revised 15 August 2021; Accepted 17 August 2021; Published 20 September 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Congcong Luan and Peng Shang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the rapid development of artificial intelligence and deep learning in recent years, many universities have put forward the goal of achieving digitalization, intelligent, and education informatization on campus. Throughout the lecture and learning process, the classroom status is an important reference factor to assess students' acceptance of the course and the quality of lectures. However, at present, classroom status analysis is mainly conducted manually, which can distract teachers' attention, so it is of great research significance to find a method that can improve the efficiency of classroom status analysis. In this paper, we choose an offline method to analyze the status of a classroom video recording in terms of students' behavior and attendance in terms of frames, in which student behavior is identified by an improved target detection algorithm and attendance is analyzed by face recognition. By analyzing the structure of the neural network model, an improved neural network model is proposed for its characteristics of a large number of parameters and poor detection of small targets in the basic network. The backbone network is replaced by the improved neural network, and the depth-separable convolutional network is used to reduce the network parameters and increase the computation speed. The information in the deeper feature map is fused upward into the shallow layer to improve the accuracy of small target recognition. Finally, the optimization algorithm is incorporated into the network to optimize the network model and accelerate the model convergence speed. In addition, this paper incorporates the improved behavior recognition method and face recognition method into the system to realize the analysis of the offline classroom status. The system is divided into a teacher side and a management side, where the teacher side is responsible for uploading course recordings and the management side is responsible for randomly analyzing students' status and attendance at any time, and the combination of the two forms a convenient and comprehensive classroom status analysis system platform. Users can upload classroom videos through the instructor interface and can view the classroom status analysis results of a course at any time by searching randomly in the administration. In this paper, the classroom status is mainly judged by the recognition of students' behaviors.

## 1. Introduction

In the era of rapid development of technology, college education is also progressing, mainly in the aspects of campus life and classroom teaching [1]. It mainly includes "Intelligent Library," "Smart Campus Card," "Access Control Management," and "Smart Dormitory" [2]. Previously, you needed a meal card to eat, a library card to borrow books, a bath card to take a bath, and an access card to enter

and leave the dormitory, but the mobile smart card integrates these functions into one campus one-card [3]. Meanwhile, reforms in classroom teaching are also ongoing, and many innovative educational theories have emerged, such as flipped classroom, teaching theory, constructivism theory, and theme-based education theory. These theories have given great impetus to the teaching reform. However, there are still many problems in traditional classroom teaching [4]. At present, classroom attendance is mainly in the form



of roll call, which is fine in small classes, but not in large classes with hundreds of students, where it takes about thirty minutes to roll call everyone [5]. Not only is it time-consuming and labor-intensive, but there will also be people with answers. Another way to do roll call is to sign in via mobile app. Although this way saves time, there will be network lag when there are more people, so students cannot sign in on time; in addition, there are also cases of people taking the cell phone and signing in for them [6].

As the main subject of classroom teaching, students' classroom behavior can reflect the degree of acceptance of knowledge and will also directly affect the quality of teachers' lectures [7]. In traditional classroom teaching, teachers can only understand the status of a class through their own observation, which undoubtedly increases the workload for teachers, and the results obtained can be rather one-sided, so there is an urgent need for methods that can improve the efficiency of classroom status analysis [8]. With the development of artificial intelligence, using it to improve campus life and teaching quality has become an important direction for future research in education. A study of the development of AI reveals that it first appeared in 1943, reached its golden age in 1956-1974, and then experienced two lows and a boom. Nowadays, after a long development, it has become an interdisciplinary discipline, penetrating into all walks of life and even daily life, especially in the field of education, which has been developing rapidly in recent years [9]. On the integration of AI and education, it is pointed out that the research on the application of AI in education is at a historical inflection point, starting from the emergence of research on AI in education in 2017 to the explosive growth of related research from 2018. It is the future development trend to use innovative technologies, reform the methods used in teaching, construct a new education system containing innovative learning and interactive forms of learning, and realize intelligent teaching in accelerating the construction of the talent training model [10]. The target detection and face recognition algorithms used in this paper are part of the artificial intelligence applications. Target detection is used to locate and classify targets, for example, in intelligent driving to determine whether there are obstacles on the road and the exact location of the obstacles. The main purpose of face recognition is to identify who the detected face is and is divided into two forms: 1:1 and 1:N. Among them, 1:1 solves the problem of "whether this person is someone," such as judging whether the ID card is the person in the security check [11]. And 1:N solves the problem of "who is this person" by finding the identity of this person in a large amount of face data based on the image provided. Using the target detection method to identify student behavior, we can quickly count the number of people in the classroom for each behavior, which is convenient and accurate compared to manual methods. Using face recognition in classroom attendance, it is possible to automatically identify people who have arrived by simply entering the class image, which is fast and avoids phenomena such as proxy signing [12]. Therefore, using deep learning to analyze classroom status has practical research value and significance [13].

The training set is put into the behavior recognition network model for training to get the initial model, and the validation set is used to validate the model and adjust the network model parameters according to the validation results. The test data is put into the model to get the results, and then the results are analyzed to determine if they are the same as what is expected. Through the analysis of the previous contents, we learned that the relevant techniques and theories have been developed more and are more advanced, but not many of them have been applied to students' classroom states. Therefore, in this paper, we review the literature on target detection technology and face recognition technology, study the application methods of both in other fields, and improve or migrate the two algorithms according to the characteristics of students' classroom status to make them suitable for this study. In this paper, we choose a model combining the MTCNN face detection algorithm in pyramid mode and InsightFace face comparison algorithm for face recognition, which can meet the effect of real-time recognition and has a high accuracy rate in practice. The system is designed and implemented using the Flask+Vue framework, divided into two main parts: the teacher side and the management side. The teacher side is the front-end of the system responsible for uploading classroom video recordings to the database and viewing student status analysis and attendance results, while the management side is the back-end of the system responsible for personal management, comprehensive query, status statistics, attendance management, teaching resources, and management of basic resources.

## 2. Related Work

The traditional algorithm for detection of targets can be summarized in three main steps: first, selecting candidate regions in the image; second, extracting the features in the candidate regions where they are located; and finally, recognizing them using a classifier in the last step [14]. The Harr-like features proposed by Jakonen et al. to describe the face, also known as Harr features, can be divided into linear features, edge features, point features, and diagonal features; they proposed a method to obtain the features of the directional gradient histogram, by which the features are obtained. The way to get features is to count and calculate the gradient histogram for a certain range of the image [15]. The method of manually designed features has some drawbacks, such as the problem that the image features are not obvious and difficult to set, the designed features cannot be used under certain conditions, and the robustness is not good [16]. The AdaBoost method was proposed by Gardner, who automatically selected some weak classifiers with low recognition accuracy and combined them to increase the judgment [17]. Support vector machines achieve binary classification of data by supervised learning and have higher accuracy on linearly separable datasets, often by introducing kernel functions to map a low latitude to a high latitude to make the data linearly separable.

Decision tree is an inductive learning method based on examples, which distills a tree classification structure in

unordered data and then makes classification judgments. Random forest consists of many decision trees and classifiers, and the plurality of all the trees contained determines their output results, and the method will improve the classification accuracy [18]. Deep learning-based target detection is gradually becoming mainstream, and its detection accuracy is much higher than the traditional one. Researchers combined the RCNN algorithm and the SPP-net algorithm to propose the Fast R-CNN algorithm, which incorporates object recognition and position correction into a single network for training to reduce memory usage. The Faster R-CNN algorithm is based on the Fast R-CNN with the addition of a region generating network (RPN), which increases the speed of the algorithm [19]. The above-mentioned algorithm is called a two-stage algorithm, which classifies the candidate frames first and then is not fast enough although it has high accuracy [20]. Therefore, researchers first proposed the one-stage YOLO algorithm, which combines the generation of candidate frames and classification together, and the recognition accuracy is worse than that of Faster R-CNN but faster. This method is comparable to Faster R-CNN in terms of accuracy but much faster than it [21].

In the field of target detection, researchers have conducted a lot of research and achieved good results. The researchers improved the algorithm of YOLO V3 [22] by increasing the size of the feature map directly, instead of increasing a portion of it and then combining it with the residual blocks [23] to improve the detection of small targets. The researchers proposed a small target detection algorithm based on Faster-RCNN combined with multiscale feature fusion [24] and online hard case mining for airport scenes to improve the accuracy of small target object detection in airport scenes. Computer vision [25–27] development goal is to make it gradually close to the human eye function or even better than the human eye to complete some difficult tasks. For example, for the most applied face recognition now, the technology not only has the identity recognition function outside like human eyes but also is much faster, which can quickly compare from a large amount of data to find the target face. The development and application of computer vision and surveillance facilities are inseparable, and the combination of the two has been widely used in intelligent surveillance, real-time patient monitoring [28], virtual reality, intelligent robotics, etc. Pierce divided the computer vision based on video surveillance into three directions in his paper: behavior recognition [29] and analysis, tracking detection, and motion target detection, and proposed that behavior recognition analysis is the main direction of future development. A search on the Internet for the keyword “behavior recognition” yields 8112 results, which shows that the research on it is very high. By reading the relevant literature, it is concluded that the development of behavior recognition can be divided into two major directions: behavior recognition with artificially selected features and behavior recognition based on deep learning [30–32], and each direction can be divided into more detailed categories according to different research methods.

### 3. Neural Network Recognition Model for Classroom Interaction Benchmark Map

*3.1. Construction of Classroom Behavior Recognition Model.* The construction of the model mainly includes determining the network structure, preparing training data, and training and testing the model, which leads to the design flow of the behavior recognition model as shown in Figure 1.

The first step was to prepare images of student behaviors. We collected 2500 images of each of the five student behaviors: raising hands, sitting, writing, sleeping, and playing with a cell phone, and 500 images of each behavior. The second step was to build a behavior recognition database. The collected 2500 images were preprocessed and labeled, and the images were proportionally divided into three parts: training set, test set, and validation set. Finally, the model is trained and tested. The training set is put into the behavior recognition network model for training to get the initial model, and the validation set is used to validate the model and adjust the network model parameters according to the validation results. The test data is put into the model to get the results, and then, the results are analyzed to determine if they are the same as what is expected. We decide whether to continue training the model based on the comparison results. We save the behavior recognition models with better recognition results for subsequent classroom behavior recognition.

*3.2. SSD Target Detection Algorithm.* The target detection algorithm used in this paper is based on the improvement of the traditional SSD target detection algorithm, so the structure of the traditional SSD algorithm model and the principle of the algorithm are briefly explained before introducing the improved algorithm. SSD algorithm, a regression-based target detection model, was proposed in 2016. Depending on the input image size, SSD is divided into SSD300 and SSD512, and SSD300 is used in this paper. Its network structure is divided into two parts: first is the main part of the network, which is generally called the base network, derived from some subtype networks; second is the convolutional network added subsequently, which is used to help the previous network to further extract image features. The last fully connected layer of VGG16 is removed, and only the previous convolutional network is used, and the two new convolutional layers, named Conv6 and Conv7, are placed in the place of the just-deleted fully connected layer, and finally, eight decreasing convolutional layers are added at the end, followed by the classification layer and the nonmaximum suppression layer, as shown in Figure 2.

According to the previous section, SSD and YOLO are both one-stage-type target detection algorithms, but they are different in terms of feature extraction. The early YOLO algorithm only extracts the information of the top-level features by convolution operation, which is semantic but may lose the information of small targets, so as mentioned above, the early YOLO algorithm is fast but the detection rate of small targets is not high. The SSD algorithm uses multiple scales of feature maps for detection and adds gradually decreasing convolutional layers after the modified VGG16 base network and then selects six layers for prediction from

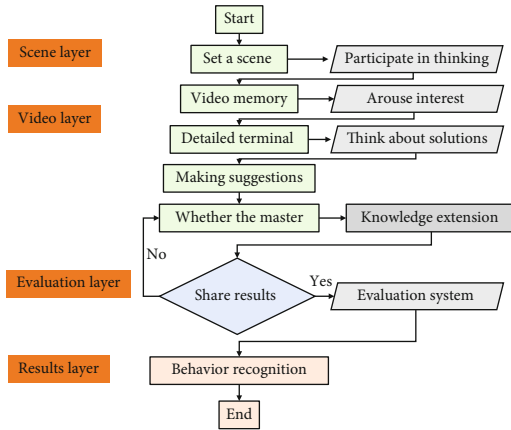


FIGURE 1: Classroom behavior recognition process.

all the layers, which are Conv4\_3, Conv7, Conv8\_2, Conv9\_2, Conv10\_2, and Conv11\_2, and their sizes are from front to back. The feature map size is gradually effective, where the larger ones are used to identify the smaller objects and the smaller ones are used to identify the larger objects. The SSD algorithm mimics the Anchor mechanism in the Faster R-CNN, but the two are very different in the application of Anchor. Faster R-CNN is a classical two-stage target detection algorithm consisting of Fast R-CNN and RPN, where the core of RPN is the Anchor mechanism responsible for generating target regions. The detection is done in two parts: first, the most likely regions are identified by a network dedicated to selecting regions, and then, the objects in the regions are classified by a target classification network. This completes the category prediction. In addition, unlike Faster R-CNN, which only uses Anchor in the last layer to generate candidate regions, SSD uses the Anchor mechanism in all the six different size feature maps mentioned above to achieve the purpose of multiscale detection.

**3.3. Topology Improvement of Neural Network.** From the introduction of the principle of the SSD algorithm above, it can be seen that the SSD algorithm mainly extracts image features through the underlying network and additional convolutional layers and then selects some feature layers for the target detection work. Although this algorithm has achieved good results in the field of target detection, there are still some shortcomings. The traditional SSD algorithm uses a modified VGG16 as the base network, which has a good classification effect but has a large number of parameters [33], i.e., 14,122,995 parameters excluding the final fully connected layer, and about 4/5 of the training time is spent on the base network, so it is difficult to train and requires a high computer configuration, and the real-time performance is not good enough. In addition, the network structure of the SSD algorithm shows that the detection of small targets is done at the shallow level of the feature map, but this level contains less information about the features, so the detection effect is not good enough. In this section, we propose the following improvement strategies for both the basic network and the small target detection, replacing VGG16 with a lightweight network to improve the detection

speed by reducing the number of parameters, and fusing the high-level semantics to the lower level to improve the small target detection. The improvement principle and the process are described in detail in the following. The goal of the basic network improvement is to replace the original backbone network VGG16 with a lightweight network. It uses deep separable convolution instead of normal convolution to reduce the number of parameters, and Mobilenet has only 4.2 million parameters compared with the 133 million parameters in VGG16. The results of both tests on the ImageNet dataset show that Mobilenet is much faster and the accuracy is only 0.9 percentage points lower than that of VGG16. Therefore, this paper uses the original Mobilenet as the base network for SSD with some modifications. The Mobilenet improvement and base network replacement process are described in detail as follows.

The topological network is faster and less computationally intensive than, for example, VGG16 because it differs in two ways: firstly, the network is composed with depth-separable convolution, and on the other hand, the width and resolution coefficients are also used. The main part is the depth-separable convolution, which completes a convolution operation in two parts: depth convolution and point convolution. The Mobilenet network structure has 28 layers if both are considered as two layers and 14 layers if they are considered as one layer. As mentioned above, the depth-separable convolution means that the convolution operation is implemented in two steps. When the image is input to the network, it needs to go through the deep convolution operation to get some maps containing the feature information, then the BN and ReLU operation to get some other feature map information after the feature map obtained above, and then the BN and ReLU operation again to get the result. The process is shown in Figure 3, with depth convolution on the top and point convolution on the bottom.

In order to reduce the network parameters, in addition to the depth-separable convolution, the width factor  $\alpha$  and the resolution factor  $\rho$  are used, with values between 0 and 1. The most common values of  $\alpha$  are 1, 0.75, 0.5, and 0.25.  $\alpha$  serves to reduce the number of channels, for example, for an input channel with a value of  $R$ , when added, it becomes  $\alpha R$ , which reduces the computational effort by about  $\alpha 2$ . Another factor that affects the amount of computation is the resolution, so  $\rho$  is used to reduce the resolution of the image, and the use of this factor reduces the amount of computation of pixel values by  $\rho 2$  similar to the reduction of  $\alpha$ .

## 4. Results and Analysis

Big data provides data support for exploring personalized group teaching models, which helps to effectively promote the transformation of group teaching to individualized student learning. Teachers can use high-end big data analysis technology, which can focus on the overall microlearning performance of students at each stage in real time, such as student response scores for each multiple-choice question and long-term student learning performance in school. The big data learning platform will provide school teachers with timely, authentic, and unique teaching information in order

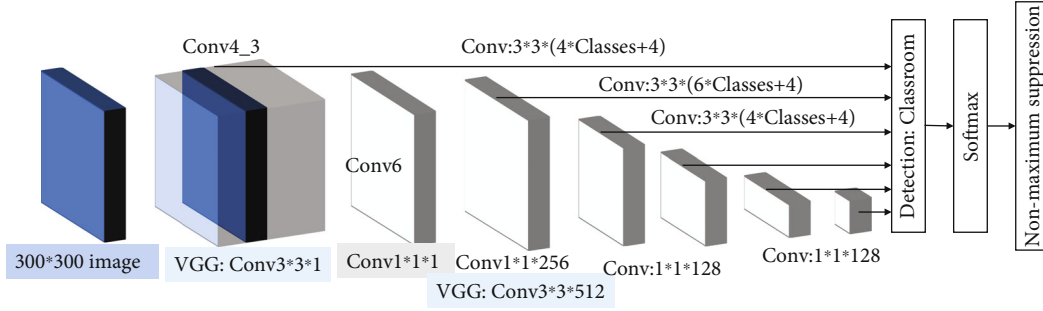


FIGURE 2: SSD network structure.

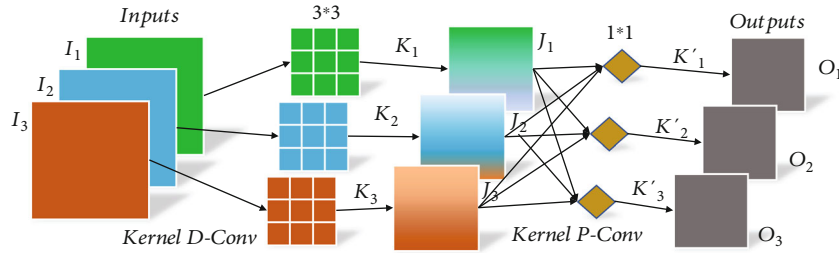


FIGURE 3: Deeply separable convolution operations.

to help teachers tailor their teaching in the classroom management process—which stages of the learning management process in the teacher’s classroom should students pay special attention to when reviewing the basics section, which stages should students pay special attention to when incorporating the practical class, at which stage should students pay special attention for the review of the basics, at which stage should they pay special attention for the integration of the contents of practical classes, at which stage should they pay attention for the strengthening of comprehensive exercises, and at which stage can they concentrate on reading the books recommended by the teacher, etc. In addition, after successfully completing the mathematics assignments assigned by the school teachers, students may also use the Smart Learning Assessment System to continuously strengthen their independent learning. The intelligent system [34, 35] is to recommend mathematics homework; if some students can answer all the recommended questions of a certain type several times correctly, they can naturally skip similar recommended questions; on the contrary, the effect is more intensive, so that not only can it greatly improve the efficiency of the candidates in learning but also it can greatly reduce the burden of some students in later learning.

Through the data collection of daily assignments and exams, we can understand the common problems and individual problems of the class, and the teacher can explore the shortcomings of the students through the data analysis. In the process of data interpretation, the teacher can compare the average score of each class, analyze the gap of each question, and find the common problems and individual problems of the class, and the teacher can strengthen the training of students for the common problems. Individual problems can be instructed individually. There may be some

obstacles between the information of the questions in the test paper or homework and the students’ goal achievement, and these obstacles can be found through the data analysis of the big data products of the extreme class, which is difficult to dig based on manual collection. The high-frequency errors and error-prone questions of the student population are also a reflection of their lack of knowledge and problem-solving skills. As academic data is collected, each student’s level in the academic diagnostic evaluation is automatically divided. Combining the students’ levels and scores, we can identify common problems and focus on cultivating excellence and transforming late-comers. With the collection of academic data, each student’s “Classmates” account will accumulate the wrong questions in the subject, forming an electronic diagnostic academic file for each student, which can be revised on cell phones and tablets with “Classmates,” and can also be exported and printed into a book. This is an invaluable resource for students and provides a solid foundation for adaptive learning research and personalized tutoring.

In this paper, we evaluate the model in terms of single-frame image detection time and mean average precision ( $P$ ) of image detection. AP is the average of all class  $P$  values. AP is the area under the line of the curve consisting of precision and recall.

The accuracy formula is as follows:

$$P = \frac{Tp + Fp}{(Tp^2 + Fp^2)^{1/2}}, \quad (1)$$

where  $Tp$  means the number of samples for which the classifier classifies the target as a positive sample and is

actually a positive sample as well. Fp refers to the number of samples that the classifier considers as positive samples but is actually negative samples.

The whole equation represents the proportion of positive samples that the classifier considers as positive samples and is indeed a positive sample itself to the overall positive samples identified by the classifier, reflecting the checking function of the model.

The formula for the recall rate is as follows:

$$\text{Recall} = \frac{\text{Tp}^{1/2}}{(\text{Tp}^2 + \text{Fp}^2)^{1/2}}, \quad (2)$$

where TN refers to the number of samples for which the classifier considers the target as negative samples but is actually positive; the whole equation indicates the proportion of samples that the classifier considers positive classes and is indeed positive classes to all positive classes, reflecting the check-all function of the model.

The squared error criterion is generally used, and the relevant definition is as follows:

$$E(n) = \iint (\lambda t)^{n+1} e^{\lambda t}. \quad (3)$$

The algorithm uses probability-based distances as a measure function:

$$d(k, n) = (\lambda t)^{n+1} e^{\lambda t} + \xi_i, \quad (4)$$

$$\zeta_n = H_{N,v} \left( \sum_{i=1}^k \sum_{P \in Q} (\lambda t)^{n+1} e^{\lambda t} + Q_n \right),$$

$$\tilde{E}(n) = \sum_{i=1}^k \sum_{P \in Q} (\lambda t)^{n+1} e^{\lambda t} \log_n \frac{N_v}{N_n}.$$

The traditional SSD algorithm, Mobilenet-SSD, and Mobilenet-SSD with feature fusion are trained under the same experimental environment and parameters, and then, the three algorithms are compared by a test set. The test environment is shown in Figure 4, and the data used for the test are the homemade datasets in this paper. Figure 4 shows the average accuracy and detection speed (detection time per image frame) of classroom behavior detection obtained using different models on the test set.

It can be clearly seen from Figure 4, we can see that in the student classroom behavior recognition test, the Mobilenet-SSD with feature fusion improves the detection speed by 2 frames per second compared with the traditional SSD algorithm, and the average accuracy of detection reaches 83.08%. Compared with the Mobilenet-SSD model without feature fusion, the speed is reduced by 2.5% because the fusion increases the network parameters, but the accuracy is improved by 6.94%. After the above analysis, this algorithm has good performance compared with the other two algorithms in terms of detection speed and recognition accuracy. Comparing the loss function change curves during the training process, we can judge the training difficulty of

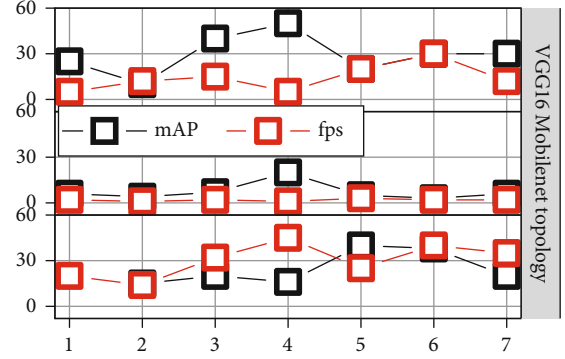


FIGURE 4: Comparison result of recognition effect of different models.

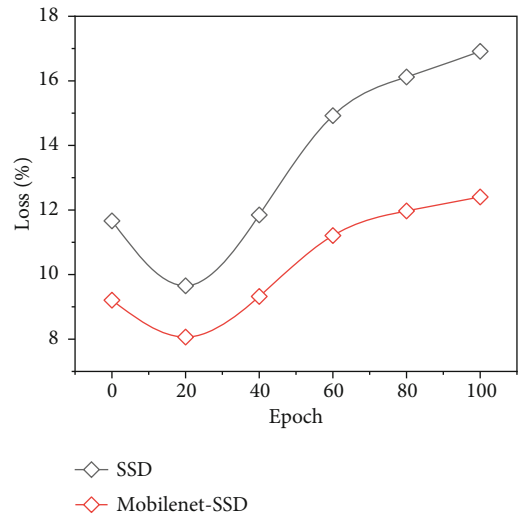


FIGURE 5: Loss variation graph.

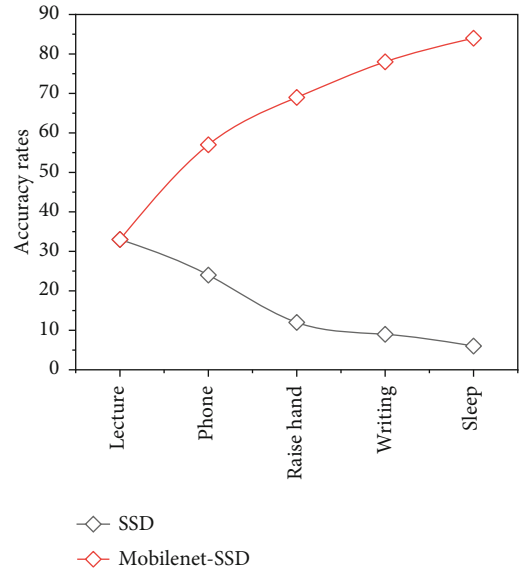


FIGURE 6: Comparison of the accuracy rates of the five classroom behavioral tests.

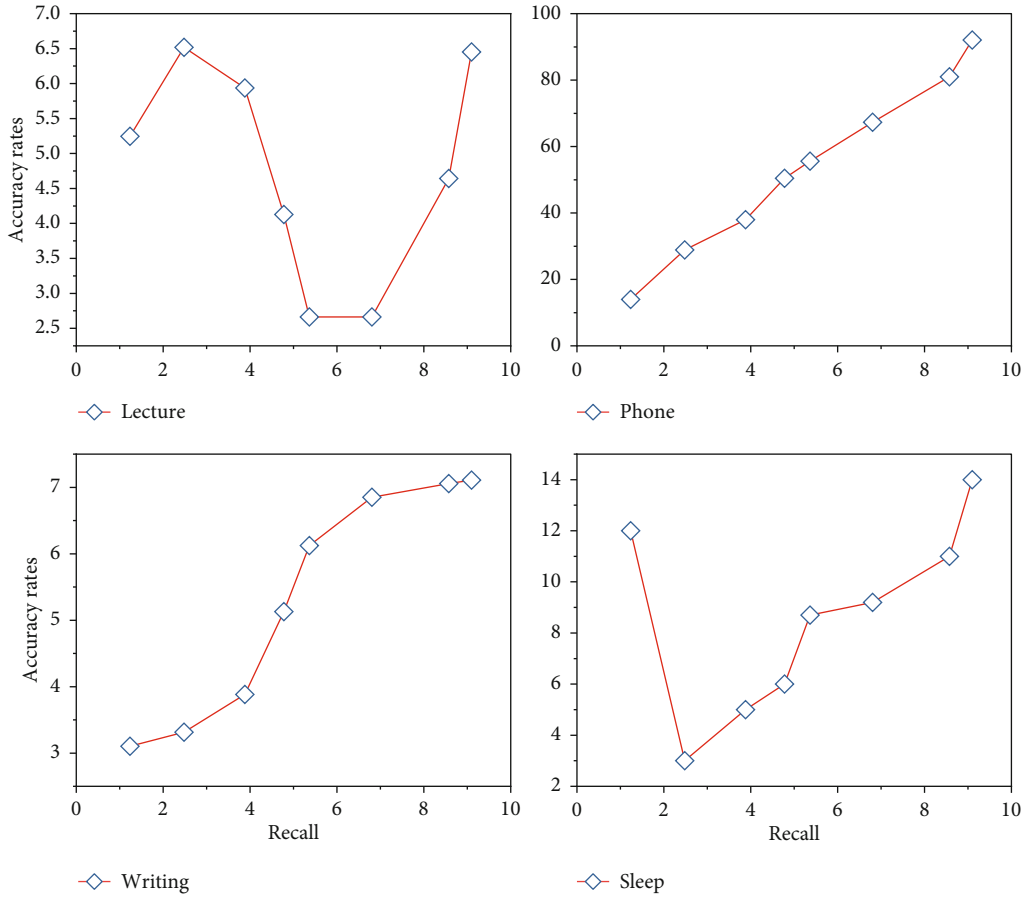


FIGURE 7: Recognition accuracy graph.

the model. The loss function curves of Mobilenet-SSD and SSD models with the same parameters and epoch of 100 for 50,000 iterations are shown in Figure 5.

It can be seen that the loss values of both models have been decreasing, which means that both models are reasonable. It took about 6 days for the loss to fall below 0.5 during training and about 8 days for the original SSD model, and it can be seen from the graph that the loss value of this model decreases faster than that of the traditional SSD model. The accuracy (AP) of each action is shown in Figure 6, which is obtained by using the Mobilenet-SSD model after SSD and feature fusion to detect five actions of students in the test set: listening, raising hands, writing, sleeping, and playing with cell phones.

From Figure 6, it is concluded that the Mobilenet-SSD algorithm after feature fusion in this paper has improved the detection of small targets in all five actions compared with the original SSD algorithm, among which writing has improved the most by 3.03%, indicating that the model in this paper has improved in small target recognition. The results of the Mobilenet-SSD recognition of feature integration showed that among the five actions, the detection accuracy of listening action was the highest and hand raising was the second highest, while writing and playing with cell phone were the lowest. The reason for this result is that the two actions are more likely to be obscured than the other

actions, especially in the recognition process, which can easily confuse other hand movements, and thus, the calibration effect is not as good as the other three actions. In order to show more intuitively the recognition accuracy of the model on the five actions, a line graph is shown in Figure 7, where the shaded area is the accuracy rate.

This paper first analyzes the classroom behavior recognition model design process and then explains how the database is constructed, including data acquisition, image preprocessing, and dataset labeling. Then, the network structure of the traditional SSD model is described, and its advantages and disadvantages are analyzed. After that, the principles related to depth-separable convolution are explained, based on which the basic network structure of Mobilenet is introduced, and the methods that are more frequently used in fusing the features of each layer of the network are analyzed. Based on the above preparatory work, an improved strategy is proposed to change the basic network to Mobilenet and to use the add method to complete the fusion of features, so that the model can improve the effect of recognizing small objects compared with the original SSD method, and the recognition speed is similar to the previous one, as shown in Figure 8.

Only the Gs plays a role, which is equivalent to a Gaussian filter that mainly acts as an image smoother. In an image, there are obvious color or light/dark shifts in the critical areas, which are reflected at the pixel level by the large

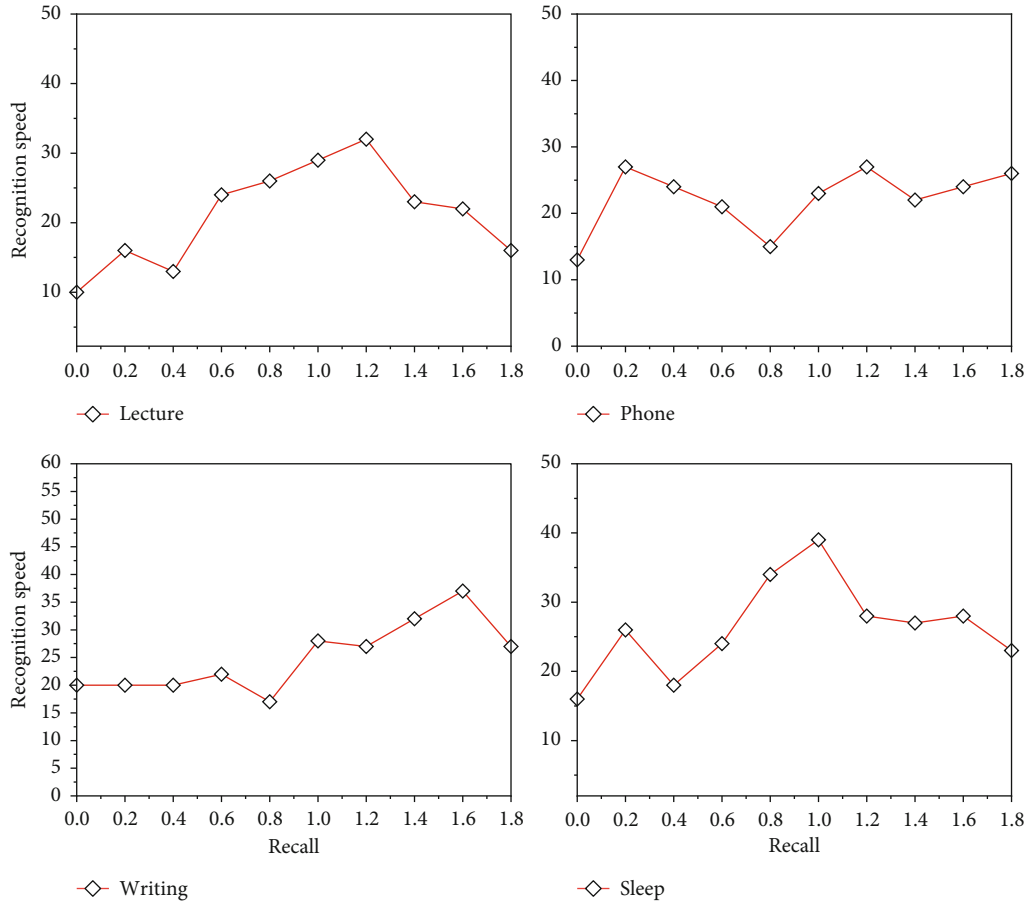


FIGURE 8: Recognition speed comparison.

difference in pixel values between the two sides. In this case, the Gr value is close to 0, and the whole filter result is 0. At this point in the whole filtering process, the value there does not affect the overall output value, which plays a role in protecting the edge effect. Image target enhancement [36] is also called sharpening, using methods such as Sobel (Sobel) processing method, USM algorithm, Laplace (Laplace) processing method, the first-order form of Prewitt processing method, and the canny processing method involving multiple orders; by comparison, this paper uses USM (Unsharpen Mask) sharpening enhancement algorithm for target enhancement. For the management, it is important to keep abreast of the teaching and learning process. For the management, it is more timely to grasp the teaching situation. What teachers do is conscientious work, and real data makes teachers' work results more tangible and transparent. In the process of teaching management implementation follow-up, you can clearly understand who is doing well and who is not doing well, so as to praise teachers for excellence in time and remind teachers for deficiencies in time. The basic theoretical principle of teaching is to teach according to the material, and the basic premise of teaching according to the material is to read and understand students and the important way to read and understand students which is also the academic diagnosis of teachers and students. Only through the academic diagnosis of the candidates is it possible to really read

a student, and by reading a student is it possible to really implement the teaching according to their abilities. Due to the lack of analysis of the teachers' individual teaching tracking survey, it is not conducive to the effective organization of individualized teaching and differentiated teaching of students, which ultimately does not achieve the effect of teaching according to their abilities. By continuously carrying out this personalized teaching, differentiated teaching, it may not find the growth point on the quality of teaching. To improve the examination promotion rate, you can only do personalized and differentiated teaching without extending the time, and to achieve personalized and differentiated teaching, use the extreme class big data.

## 5. Conclusion

In this paper, by changing the base network in the original SSD from VGG16 to the improved Mobilenet network, and using the add feature fusion method for the replaced network, the base network parameters are reduced while deep information is incorporated in the shallow layer, thus improving the effect and speed of small target detection. Then, the model is trained and the trained model is used to identify the student behaviors and get the distribution of the number of students in each of the five behaviors in a class, and then, we get the percentage of the number of

students in the five states of serious, good, fair, poor, and other as set in the thesis to complete the analysis of the student states of the class. This is combined with the thesis algorithm model to design and implement a college classroom status analysis system. Through daily homework and test data collection, we can understand common and individual problems in the classroom, and teachers can explore students' deficiencies through data analysis. In the process of data interpretation, teachers can compare the average scores of each class, analyze the gaps in each question, and find common and individual problems in the class. Teachers can strengthen the training of students for common problems. The web-oriented classroom status analysis system is designed and implemented by using the Flask+Vue framework and combining the behavioral recognition and face recognition algorithm parts. Users can upload classroom videos through the instructor interface and can view the classroom status analysis results of a course at any time by searching randomly in the administration. In this paper, the classroom status is mainly judged by the recognition of students' behaviors. Although the model recognition effect is ok, the reference factors are a bit one-sided, and the classroom status analysis can be done together with the recognition judgment of facial expressions in the future. Secondly, there is room for expansion of this system.

At present, it only analyzes the overall situation of a course during the class, and in the future, it can add the status analysis function for each student and generate the report of students' classroom status analysis according to the statistical results.

### Data Availability

The data used to support the findings of this study are included within the article.

### Conflicts of Interest

All the authors do not have any possible conflicts of interest.

### Acknowledgments

This study was supported by the National Natural Science Foundation of China (Grant Nos. 61675164 and 61827827).

### References

- [1] C. Chang and H.-C. K. Lin, "Classroom interaction and learning anxiety in the IRs-integrated flipped language classrooms," *The Asia-Pacific Education Researcher*, vol. 28, no. 3, pp. 193–201, 2019.
- [2] S. K. Ertesvåg, P. Sammons, and U. Blossing, "Integrating data in a complex mixed-methods classroom interaction study," *British Educational Research Journal*, vol. 47, no. 3, pp. 654–673, 2021.
- [3] M. Reed, "Classroom interaction and the development of empowerment," *Changing English*, vol. 25, no. 2, pp. 115–134, 2018.
- [4] A. M. Blaine, "Interaction and presence in the virtual classroom: an analysis of the perceptions of students and teachers in online and blended advanced placement courses," *Computers in Education*, vol. 132, no. 1, pp. 31–43, 2019.
- [5] O. Metatla, A. Thieme, E. Brulé, C. Bennett, M. Serrano, and C. Jouffrais, "Toward classroom experiences inclusive of students with disabilities," *Interactions*, vol. 26, no. 1, pp. 40–45, 2018.
- [6] A. Ogan, "Reframing classroom sensing," *Interactions*, vol. 26, no. 6, pp. 26–32, 2019.
- [7] V. Pavón Vázquez and M. C. Ramos Ordóñez, "Describing the use of the L1 in Clil: an analysis of L1 communication strategies in classroom interaction," *International Journal of Bilingual Education and Bilingualism*, vol. 22, no. 1, pp. 35–48, 2019.
- [8] K. Solheim, "Teachers' aspirations to improve their classroom interaction," *International Journal of Learning Teaching and Educational Research*, vol. 18, no. 6, pp. 147–169, 2019.
- [9] C. Bokhove, "Exploring classroom interaction with dynamic social network analysis," *International Journal of Research & Method in Education*, vol. 41, no. 1, pp. 17–37, 2018.
- [10] K. Solheim, S. K. Ertesvåg, and G. Dalhaug Berg, "How teachers can improve their classroom interaction with students: new findings from teachers themselves," *Journal of Educational Change*, vol. 19, no. 4, pp. 511–538, 2018.
- [11] J. Batlle and M. Deal, "Teacher epistemic stance as a trouble in foreign language classroom interaction," *Journal of Pragmatics*, vol. 176, pp. 15–25, 2021.
- [12] H. Gresch and M. Martens, "Teleology as a tacit dimension of teaching and learning evolution: a sociological approach to classroom interaction in science education," *Journal of Research in Science Teaching*, vol. 56, no. 3, pp. 243–269, 2019.
- [13] U. Karvonen, L. Tainio, and S. Routarinne, "Uncovering the pedagogical potential of texts: curriculum materials in classroom interaction in first language and literature education," *Learning Culture and Social Interaction*, vol. 17, pp. 38–55, 2018.
- [14] A.-M. Vogler, S. Prediger, U. Quasthoff, and V. Heller, "Students' and teachers' focus of attention in classroom interaction — subtle sources for the reproduction of social disparities," *Mathematics Education Research Journal*, vol. 30, no. 3, pp. 299–323, 2018.
- [15] T. Jakonen, "Retrospective orientation to learning activities and achievements as a resource in classroom interaction," *The Modern Language Journal*, vol. 102, no. 4, pp. 758–774, 2018.
- [16] P. L. Hartwick, "Investigating research approaches: classroom-based interaction studies in physical and virtual contexts," *ReCALL*, vol. 30, no. 2, pp. 161–176, 2018.
- [17] R. Gardner, "Classroom interaction research: the state of the art," *Research on Language and Social Interaction*, vol. 52, no. 3, pp. 212–226, 2019.
- [18] S. K. Ertesvåg and T. Havik, "Students' proactive aggressiveness, mental health problems and perceived classroom interaction," *Scandinavian Journal of Educational Research*, vol. 65, no. 1, pp. 1–20, 2021.
- [19] M. S. Lerang, S. K. Ertesvåg, and T. Havik, "Perceived classroom interaction, goal orientation and their association with social and academic learning outcomes," *Scandinavian Journal of Educational Research*, vol. 63, no. 6, pp. 913–934, 2019.
- [20] T. R. D. Roh and Y. A. Lee, "Teacher repetition as an instructional resource for classroom interaction: three pedagogical actions in kindergartens in an Efl context," *System*, vol. 74, pp. 121–137, 2018.



- [21] A. Llinares and N. Evnitskaya, "Classroom interaction in CLIL programs: offering opportunities or fostering inequalities?," *Tesol Quarterly*, vol. 55, no. 2, pp. 366–397, 2021.
- [22] Z. Huang, P. Zhang, R. Liu, and D. Li, "Immature apple detection method based on improved Yolov3," *ASP Transactions on Internet of Things*, vol. 1, no. 1, pp. 9–13, 2021.
- [23] W. Cai, Z. Wei, R. Liu, Y. Zhuang, Y. Wang, and X. Ning, "Remote sensing image recognition based on multi-attention residual fusion networks," *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 1, pp. 1–8, 2021.
- [24] C. Gao, S. Ye, H. Tian, and Y. Yan, "Multi-scale single-stage pose detection with adaptive sample training in the classroom scene," *Knowledge-Based Systems*, vol. 222, p. 107008, 2021.
- [25] J. Zhang, J. Sun, J. Wang, and X. G. Yue, "Visual object tracking based on residual network and cascaded correlation filters," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–14, 2020.
- [26] Z. Chu, M. Hu, and X. Chen, "Robotic grasp detection using a novel two-stage approach," *ASP Transactions on Internet of Things*, vol. 1, no. 1, pp. 19–29, 2021.
- [27] J. Zhang, Y. Liu, H. Liu, and J. Wang, "Learning local–global multiple correlation filters for robust visual tracking with Kalman filter redetection," *Sensors*, vol. 21, no. 4, p. 1129, 2021.
- [28] L. Yu, S. Tao, W. Gao, and L. Yu, "Self-monitoring method for improving health-related quality of life: data acquisition, monitoring, and analysis of vital signs and diet," *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 1, pp. 24–31, 2021.
- [29] X. Ning, K. Gong, W. Li, L. Zhang, X. Bai, and S. Tian, "Feature refinement and filter network for person re-identification," *IEEE Transactions on Circuits and Systems for Video Technology*, 2020, In press.
- [30] W. Chu, *Studies on the effects of wiring density on chip package interaction and design optimization with machine learning*, [Ph.D. thesis], The University of Texas at Austin, 2021.
- [31] Y. Gu, A. Chen, X. Zhang, C. Fan, K. Li, and J. Shen, "Deep learning based cell classification in imaging flow cytometer," *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 2, pp. 18–27, 2021.
- [32] W. Cai, Y. Song, and Z. Wei, "Multimodal data guided spatial feature fusion and grouping strategy for E-commerce commodity demand forecasting," *Mobile Information Systems*, vol. 2021, Article ID 5568208, 14 pages, 2021.
- [33] M. Miao, W. Cai, and X. Li, "Parameter estimation of gamma-gamma fading with generalized pointing errors in FSO systems," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 1301878, 14 pages, 2021.
- [34] Y. Jiang, X. Gu, D. Wu et al., "A novel negative-transfer-resistant fuzzy clustering model with a shared cross-domain transfer latent space and its application to brain CT image segmentation," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 18, no. 1, pp. 40–52, 2021.
- [35] A. Soller, "Supporting social interaction in an intelligent collaborative learning system," *International Journal of Artificial Intelligence in Education*, vol. 12, no. 1, pp. 40–62, 2001.
- [36] L. Zhang, X. Wang, X. Dong, L. Sun, W. Cai, and X. Ning, "Finger vein image enhancement based on guided tri-Gaussian filters," *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 1, pp. 17–23, 2021.

## Research Article

# College Oral English Teaching Reform Driven by Big Data and Deep Neural Network Technology

Hui Liu 

*Department of Foreign Language, Zhanjiang University of Science and Technology, Zhanjiang 524000, China*

Correspondence should be addressed to Hui Liu; [liuhui19811117@126.com](mailto:liuhui19811117@126.com)

Received 5 August 2021; Revised 16 August 2021; Accepted 18 August 2021; Published 18 September 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Hui Liu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The ultimate goal of English teaching is to cultivate the students' ability to communicate information in English, master good language learning methods, and become independent language learners and users. Therefore, successful English language teaching needs to be achieved through language communication training between teachers and students and between students. This article investigates the importance of promoting the reform of oral English teaching in China's English teaching environment. We believe that to promote the reform of oral English teaching, an oral teaching environment must be available. However, the current common problem in oral English teaching in colleges and universities is that the spoken conversation objects are not standard enough, or there is no person who can talk to. Therefore, an intelligent spoken dialogue system based on big data and neural network technology is particularly important, and the quality of dialogue depends on accurate spoken speech evaluation. We first extracted six features of pronunciation quality, fluency, content richness, topic relevance, grammar, and vocabulary richness. Secondly, we propose an evaluation model that connects specific TDNN layers in a feedforward manner, using the feature representation of target words in different TDNN layers, which can obtain richer context information and greatly reduce the amount of model parameters. Finally, we conducted a simulation experiment. The experimental results show that the proposed model is accurate in evaluating spoken English and can effectively assist the reform of spoken English teaching in colleges and universities, and its performance is better than SVM by 9.2%.

## 1. Introduction

In recent years, the application of information technology [1–3] in the field of education is more and more extensive. In oral English teaching, due to the increasing popularity of English teaching in China, traditional language teaching methods [4] can no longer meet people's needs, which is more and more obvious in colleges and universities. In this context, the computer-aided language learning system [5–7] based on big data and neural network [8–10] has become the focus of research. It can take the place of teachers for students' examination answers, classroom homework automation correction, so that teachers from repeated and time-consuming correction work out of the liberation. At present, such an automatic correcting system [11] has been able to achieve almost complete accuracy in the objective task. This form of question is typically used for multiple

choice, fill-in, and other types of questions, and for writing and oral questions, automatic correction is still a research topic that needs to be broken through. The oral questions are separated into two types: one is retelling, reading, and reciting known content, and the other is "open speaking," in which the exam taker provides free rein to specific questions or subjects, through the examinee pronunciation and standard pronunciation for speech level comparative analysis, with the development of speech recognition technology [12–14].

A full evaluation of candidates' answers from several dimensions, including oral fluency, rhythm, intonation, vocabulary richness, and semantics, is also necessary for open spoken English, in addition to "pronunciation correctness." For a long time, open spoken English scoring [15, 16] and evaluation technology research has not yielded significant results. With the advancement of machine learning

technology [17–19], some researchers began to investigate how to use it to automatic speech evaluation, resulting in the development of the well-known SpeechRater automatic scoring system. The system combines feature engineering and machine learning algorithms [20] to achieve automatic scoring of open spoken language, and it has also set off a wave of research in this field. Although there is still a certain gap between the scoring results of the system and the teachers' manual scoring, it provides a good research idea for later researchers. Today, new changes have taken place in the field of artificial intelligence; the most significant sign of which is the maturity of deep learning technology. Deep learning [21, 22] employs a multilayer network to do data characterization and uncover richer features. The correlation between machine and manual scores has substantially improved, and scoring errors have grown fewer and smaller after a large number of researchers applied this technology to the realization of the oral score scoring model. The open oral scoring method can now be used on a practical level thanks to deep learning technologies.

Based on the foregoing observations, we discovered that big data and deep neural network technology-driven college oral English teaching reforms [23, 24] have become a trend. Teachers will be able to dedicate more energy to actual teaching work as a result of this, and the quality of instruction will be improved. As a result, an intelligent oral dialogue system based on big data and neural network technology is critical, and the quality of discourse is dependent on precise oral evaluation. We first extracted six features of pronunciation quality, fluency, content richness, topic relevance, grammar, and vocabulary richness. Secondly, we propose an evaluation model that connects specific TDNN layers in a feedforward manner. Using the feature representations of target words in different TDNN layers, we can obtain richer context information and greatly reduce the amount of model parameters. Finally, we conducted a simulation experiment. The experimental results show that the proposed model is accurate in evaluating spoken English and can effectively assist the reform of spoken English teaching in colleges and universities.

The following are the main contributions points of this paper:

- (1) This paper reforms the college oral English teaching based on big data and deep neural network technology and proposes a spoken language recognition model, which reduces the burden of teaching and improves the quality of teaching
- (2) We propose an evaluation model that connects specific TDNN layers in a feedforward manner. Using the feature representations of target words in different TDNN layers, we can obtain richer context information and greatly reduce the amount of model parameters
- (3) We carried out an experiment with simulation. The experimental results show that the proposed model is accurate in the assessment of spoken English and can help effectively reform language teaching at universities and colleges

The following is the general structure of the paper: The background is examined in Section 2. In Section 3, some details about the suggested algorithm's concepts and related submodules are presented. The experimental results are detailed in Section 4.

## 2. Background

As a medium for people to communicate with each other, the characteristics of spoken language are very convenient and concise. It is a main communication method for people to obtain information. Nowadays, people regard whether computers can understand the spoken language used in people's daily life as a research direction of artificial intelligence, and spoken language is the natural language used by people in daily life. The oral dialogue system is a tool for people to communicate with computers. The computer understands people's spoken language and makes corresponding answers. The spoken dialogue system has been widely used in the information query system. The main reason is that the price of manual customer service is more expensive than the dialogue system, and human resources are limited. The use of this system can greatly improve the efficiency of the system, so as to serve more people. The system can greatly facilitate people's daily life and, at the same time, improve work efficiency to a large extent.

Oral English comprehension research is critical for improving the effectiveness of the oral conversation system [25, 26]. Speech recognition, oral comprehension, dialogue management, text production, and speech synthesis are the most common modules. The spoken conversation system uses a speech recognition module to transform the user's voice into text, then converts each word in the text into a corresponding word vector, and finally classifies the entire sentence or each word so that the computer can extract the phrase's main semantics. The dialog management part analyzes the user's request to get the system's answer, and then, the text generation module generates texts based on the results of the dialog management. These texts are related to the time sequence. It can be seen that oral comprehension plays a key role in the performance of the oral dialogue system.

Whether a machine allows people to carry out related tasks through the use of spoken language is the criterion for judging whether the machine is truly "smart." The dialogue system conforms to the habit of humans using spoken language for interaction. Compared with traditional information acquisition methods, the dialogue system has great advantages. The user and the system can use multiple rounds of dialogue, such as inquiries, clarifications, and confirmations, to achieve the needs of information acquisition and emotional comfort in complex scenarios. With the continuous development of science and technology in the future, robots in related fields such as service, social networking, and industry will become new members of the future society, and human-machine dialogue technology is extremely critical for whether humans and machines can achieve "intelligence." With the development of artificial intelligence technology, the human-machine dialogue

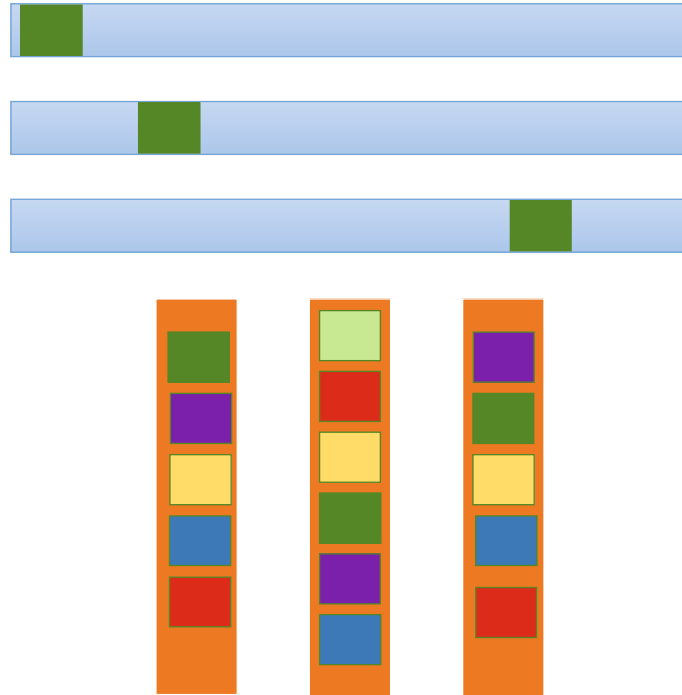


FIGURE 1: One-hot encoding and word embedding represent the word vector.

system will play an extremely important role in the future intelligent society.

### 3. Methodology

This section will elaborate on the college oral English teaching reform algorithm driven by big data and deep neural network technology. The core link is the evaluation of spoken English. This section will conduct a detailed analysis of natural language processing and neural networks.

**3.1. Natural Language Processing.** In the spoken language scoring system studied in this article, a neural network model needs to be used to score the candidates' spoken language content. The neural network cannot directly process text data, so it is necessary to convert the text into numerical data that the neural network can recognize. A commonly used method is one-hot encoding, which associates each word with a numeric vector of length  $N$ . The numeric vector corresponding to each word has only one element of 1, and the rest of the elements are 0. For example, the one-hot codes of the three words "me," "he," and "she" are  $[1,0,0]$ ,  $[0,1,0]$ , and  $[0,0,1]$ , respectively. Although this method is simple, it has the following two main disadvantages: (1) Since there is only one bit in each numeric vector to identify a word; if there are  $N$  words in the text, an  $N$ -dimensional vector needs to be used to encode it. Therefore, when the number of nonrepeated words in the text is large, the dimensionality of the vector will be large. At the same time, the number of nodes in the neural network will increase, and the calculation will become more complicated. (2) One-hot encoding, a simple encoding method, cannot describe the semantic relationship between words and thus cannot pro-

vide more information for subsequent neural network calculations. What a useful information. In order to solve the above problem, word embedding appears, which can use lower-dimensional vectors to represent words. At the same time, for words with similar meanings, their vector representations are also similar. As shown in Figure 1, the word vectors are represented by one-hot encoding and word embedding, respectively. It can be found that the latter can embed richer information into lower-dimensional vectors.

The network topology presented in Figure 2 is commonly employed when utilizing neural networks to handle natural language challenges. The word embedding layer is the network's first layer, and it turns the words in the input text into a word vector representation, such as a word with 20 letters. If the length of the word embedding vector is 50, the text of a word will become a 2-dimensional matrix of 2050 after the word embedding layer.

**3.2. Recurrent Neural Network.** The recurrent neural network (RNN) is a special artificial neural network, which is different from the feedforward neural network of the general structure. It is a neural network with internal loops. This structure enables information to circulate in the network, so unlike CNN and other networks, their output only considers the impact of the previous input and does not consider the impact of the input at other times. In RNN, the output is at every time linked not just to the input, but to the input in the preceding moment, just like the "memory" function is available in the network. Consequently, RNN is very suited for serial data processing, particularly text data. If the state is regarded at every moment as a layer of the feedforward neural network, then the cyclical network can be seen as a feedforward neural network that shares weight

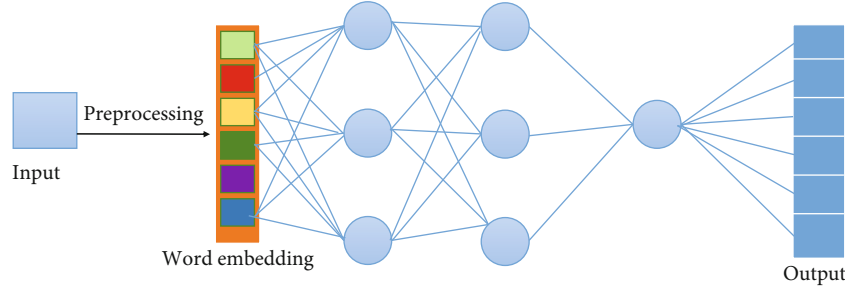


FIGURE 2: Schematic diagram of the neural network model architecture with word embedding layer.

in time. Figure 3 shows the expanded structure of a single-layer RNN network, where  $t$  is time,  $x_t$  represents the input at time  $t$ ,  $h_t$  is the state of the hidden layer at time  $t$ , and  $o_t$  represents the output of the RNN network at time  $t$ ; matrix  $W$ ,  $V$ ,  $U$  represents the weight matrix, and the calculation equations of  $h_t$  and  $o_t$  are as follows:

$$h_t = \text{func}(Uh_{t-1} + Wx_t + b), \quad (1)$$

$$o_t = Vh_t. \quad (2)$$

Although the traditional RNN network can process sequence data, it has a more serious problem: when the input sequence is long, there will be a problem of gradient disappearance during the error back propagation process, so that the network will eventually become unable to train. Therefore, the traditional RNN model is only suitable for processing short-sequence data. In order to solve the problem of insufficient “long-term memory” ability of the traditional RNN network, many researchers began to explore how to improve the model. Hochreiter and Schmidhuber proposed LSTM in 1997. This model solves the above problems well. The LSTM network introduces a new state  $c_t$  (also called a memory unit) internally for the circular transmission of information.

At each time  $t$ ,  $c_t$  records the historical information up to the current time. The state  $h_t$  of the hidden layer and the state  $c_t$  of the memory unit. The calculation equation is as follows:

$$c_t^* = \tanh(W_c x_t + U_c h_{t-1} + b_c), \quad (3)$$

$$c_t = f_t \circ c_{t-1} + i_t \circ c_t^*, \quad (4)$$

$$h_t = o_t \circ \tanh(c_t), \quad (5)$$

where  $f_t$ ,  $i_t$ , and  $o_t$  are three gate controllers, and the gate control mechanism is a method to allow information to pass through selectively. The value range of the door controller is between 0 and 1, which means that a certain proportion of information is allowed to pass.  $\circ$  represents the matrix dot product operation,  $c_{t-1}$  is the state of the memory unit at the previous moment, and  $c_t^*$  represents the candidate state of  $c_t$ .

**3.3. TDNN Model.** The TDNN algorithm is similar to the standard back spread algorithm in terms of training, and it is a quick algorithm. The TDNN is a multilayered network

with abstract capacity at each layer and the ability to achieve input sequence in time. TDNN is time invariant, and the network learning process does not necessitate precise input data placement. The benefit of TDNN is that each layer’s TDNN shares weights, making the model easier to train. The time range of the context of the sequence collected by TDNN becomes wider and wider as information flows to higher layers of TDNN. As a result, TDNN has a distinct edge in certain jobs where context information is critical.

In the phoneme recognition paper, a time delay neural network is first proposed. It is a neural feedback network of multilayered systems. The neural network time delay is a hierarchical neural feedforward network. The network can take the proximity data of the current frame into account and complete the function extraction of the current frame. The time delay neural network can be multilayered; each layer has a strong ability to abstract features; it can represent the relationship between features in time; with time invariance, it is more convenient to learn by sharing weights. In the process of learning, it is not required to carry out precise time positioning of the learned marks. The training method of the time delay neural network is the traditional back propagation algorithm.

The time and space complexity of TDNN is the same as that of the convolutional neural network. The time complexity of a single convolutional layer is

$$\text{Time} \sim O(M^2 \times K^2 \times C_{\text{in}} \times C_{\text{out}}), \quad (6)$$

where  $M$  is the side length of each convolution kernel’s output feature map and  $K$  is the side length of each convolution kernel, i.e., the number of output channels of the network’s upper layer. The output feature map area  $M^2$  determines the time complexity of each convolutional layer, as shown in the formula, input  $C_{\text{in}}$ , output  $C_{\text{out}}$ , and the convolution kernel  $K^2$ . Three parameters determine the size of the output feature map: the size of the input matrix  $X$ , the size of the convolution kernel  $K$ , padding, and stride. The following is the calculation formula:

$$M = \frac{(X - K + 2 * \text{Padding})}{\text{Stride} + 1}. \quad (7)$$

The convolutional network’s overall time complexity,  $D$ , is the number of convolutional layers in the network model;  $\ell$  is the network’s  $\ell$ th convolutional layer; and  $C$  is the

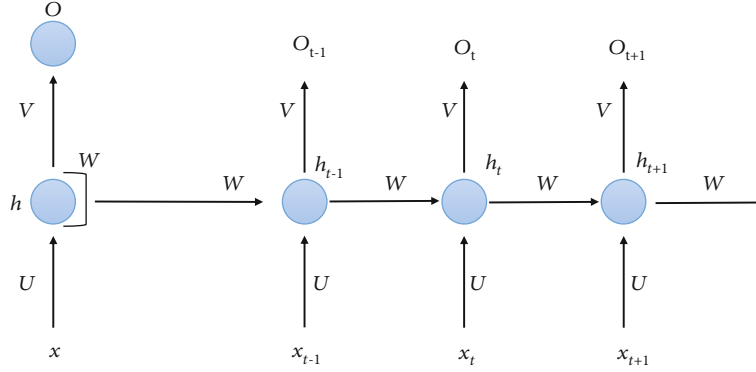


FIGURE 3: Schematic diagram of a single-layer RNN structure.

number of output channels of the  $\ell$ th convolution kernel, which is also the current convolutional layer. The number of output channels of the convolutional layer of the  $\ell - 1$  layer is the input channel  $C$  for the  $\ell$ th convolutional layer. It can be seen that the convolutional network's time complexity is the sum of the time complexity of all convolutional layers.

$$\text{Time} \sim O\left(\sum_{\ell=1}^D M_{\ell}^2 \times K_{\ell}^2 \times C_{\ell-1} \times C_{\ell}\right). \quad (8)$$

The space complexity of the convolutional network is described as the parameter quantity of the model, which is expressed as the size of the model.

$$\text{Space} \sim O\left(\sum_{\ell} K_{\ell}^2 \times C_{\ell-1} \times C_{\ell}\right). \quad (9)$$

The size  $K$  of the convolution kernel, the number of channels  $C$ , and the number of network layers  $D$  of the model are the only variables that influence the model's space complexity. The size of the input data has no bearing on the complexity of space.

**3.4. Our Model.** On the basis of multilayer TDNN, this paper proposes a parallel structure of the TDNN network (as shown in Figure 4). Multiple layers of time-delayed neural networks can be stacked to obtain more contextual information, but this causes gradient explosion and gradient dispersion issues. In response to this issue, the residual convolutional neural network has performed well on image classification tasks, demonstrating that the residual structure can reduce gradient dispersion or explosion by using jump connections. In this paper, by quoting the residual structure, the number of network layers of the model can be deepened, and its performance on the task of image classification has been significantly improved. We compared the experimental results of multilayer time-delayed neural networks and discovered that increasing the number of layers does not improve the model's performance, but rather decreases it. As a result, we add a residual structure to the multilayer time delay neural network model in order to improve the model.

The context information of the current word is well captured by TDNN, and the longer contextual information is captured by stacked multilayered delayed neural networks. The residual structure can combine the low-level network's features with the high-level network's features to improve feature representation.

In the word embedding layer, in many NLP tasks, a common practice is to construct a dictionary of words in the dataset. Each id of the dictionary corresponds to a specific word, and the id of each input word is converted to a  $D$ -dimensional real value vector, which is called a word vector. We splice consecutive  $W$  word vectors as the representation of the word vector of the current word, and  $W$  is the size of the spliced word window.  $w$  is the offset of the spliced context,  $W = 2w + 1$ . During splicing, we fill in the embedding representation of the filling symbol if there are not enough words before or after the target word. As a result, the input at  $t$  in the sequence is

$$E_t = [e_{t-w}, \dots, e_{t-1}, e_t, e_{t+1}, \dots, e_{t+w}]. \quad (10)$$

The vector of the entire sentence can be represented as an input matrix  $s \in R^{N \times W \times D}$  for a sentence containing  $N$  single words.

$$E_t = [e_{t-w}, \dots, e_{t-1}, e_t, e_{t+1}, \dots, e_{t+w}]. \quad (11)$$

In addition, the normalized probability distribution is obtained using the SoftMax activation function at the network's last layer, and the cross-entropy-based objective function used in this article is as follows:

$$L = -\frac{1}{N} \sum_{t=1}^N \sum_{c=1}^C y_{t,c} \log y'_{t,c}, \quad (12)$$

where  $y'_{t,c}$  is the true probability distribution of the  $c$ th label of the  $t$ th word in the sample and  $y_{t,c}$  is the probability of the  $c$ th label of the  $t$ th word.  $N$  denotes the number of words in the sample, while  $C$  denotes the number of semantic categories.

When it comes to predicting semantic labels for oral English comprehension, the target word's background

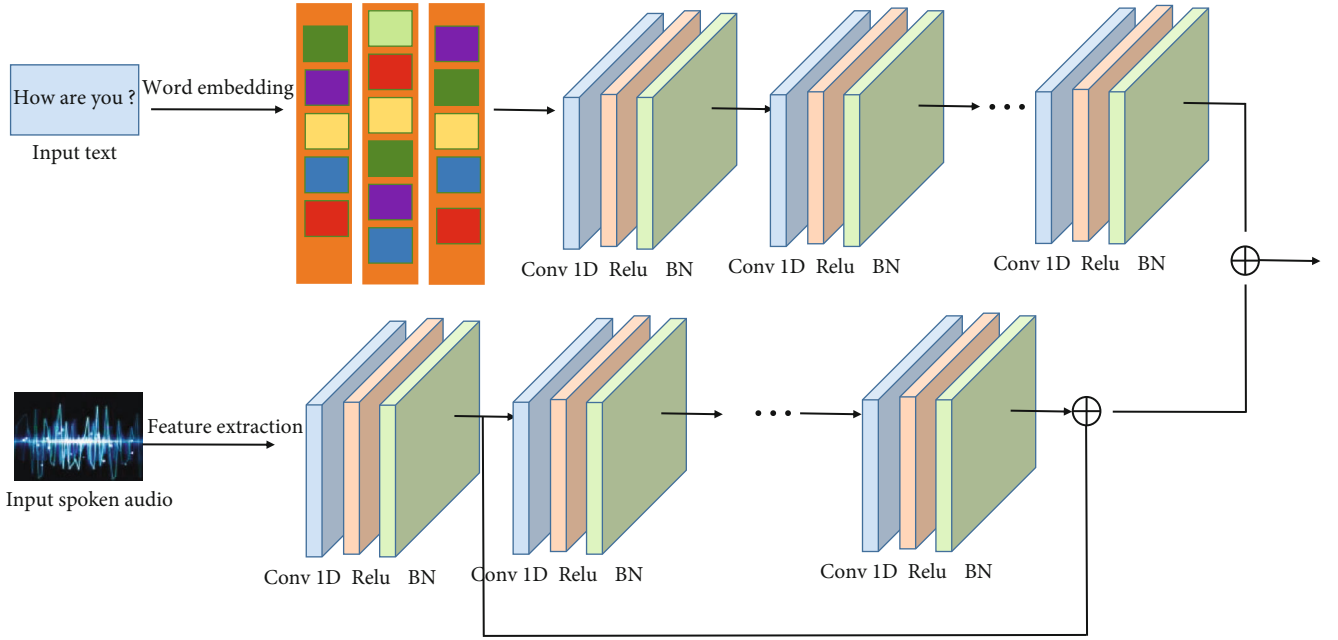


FIGURE 4: Schematic diagram of the proposed spoken English evaluation model.

information is crucial. It is vital to collect the target word's context information because the same term can have multiple labels in different circumstances. The delay offset of each layer of a delay neural network can be adjusted, changing the range of background information that the extracted features can acquire. The proposed model can successfully extract the target word's contextual information through delay migration.

## 4. Experiments and Results

**4.1. Experimental Environment.** The experiment uses the Linux operating system Centos 6.5 version and uses the deep learning tool Pytorch 0.4 version in the Python 3.6.5 environment under Anaconda. Pytorch is widely used in the field of deep learning; its code is simple and easy to write and can run on CPU and GPU. All the networks in this article are implemented using the Pytorch deep learning toolkit under the Python 3.6.5 environment. The learning rate is 0.01, and we batch processed 1,000 samples each time.

**4.2. Dataset.** The data used in this article is derived from the oral examination of a situational English course at a university. We extracted 650 test takers' answers to the same open-ended oral question from the examination data from 2012 to 2018 (each recording is about 60 seconds or so) and the teacher's manual scoring data (10-point system, including decimals). The system designed in this paper needs to score the spoken language pronunciation and the spoken language content separately, so we also asked the teacher to score the examinees' spoken language separately from these two aspects. In addition, before training and testing the model, we also need to convert the recording format. The recording files collected from the oral test are all in mp3 format, and the audio attributes are 16 bit and 16kHz sampling rate.

FFmpeg is an open source tool that specializes in processing video and audio streams. We use this tool to convert recordings in mp3 format to pcm format. Finally, we divide the entire dataset into two groups for training and testing: the "training set" contains 500 pieces of data, and the "test set" contains 150 pieces of data.

**4.3. Evaluation Index.** The evaluation criteria used in this article are F1, precision, and recall, and the calculation equations are as follows:

$$F1 = \frac{2 \times P \times R}{P + R} \times 100\%, \quad (13)$$

$$P = \frac{TP}{P_{\text{all}}} \times 100\%, \quad (14)$$

$$\text{Recall} = \frac{TP}{T_{\text{all}}}. \quad (15)$$

**4.4. Experimental Results.** In order to prove the effectiveness of the proposed algorithm, we compared SVM and BP neural network, and the comparative experimental results are shown in Table 1.

As can be seen from Table 1, the proposed algorithm achieves competitive results. Compared with SVM and BP algorithm, the accuracy of this paper is improved by 9.2% and 5.4%, respectively, and that of F1 is improved by 7.5% and 6.4%, respectively, indicating that the proposed algorithm is effective.

**4.5. Ablation Experiments.** The slot value filling results can be seen from Table 2 using the neural network of multilayer delays. The experimental delay  $D$  settings, the number of convolutionary kernels, and the size of the word concatenation window  $W$  are identical to those of the single-layer

TABLE 1: Compare experimental results.

Methods	$P$	$R$	F1
SVM	0.8102	0.8207	0.8514
BP	0.8435	0.8525	0.8624
Ours	0.8925	0.9017	0.9214

TABLE 2: Results of ablation experiments.

Layers	$P$	$R$	F1
1	0.8425	0.8936	0.9125
2	0.8169	0.8814	0.9111
3	0.8925	0.9017	0.9214
4	0.8755	0.8936	0.9125
5	0.8852	0.8745	0.9005

TDNN model to deliver best performance. To observe the effect of the TDNN network layers as a result of slot filling, just stack TDNN layers from 1 to 5. As shown in the table, the F1, with 3-layer TDNN, amounted to 92.14%. The F1 value obviously drops to 90.05% by increasing the number of network layers from 3 to 5. The experimental results show that the number of layers of TDNN could simply increase and that dependency relationship cannot be captured more effectively. It is not only difficult to train the profound network structure model but also can cause gradient problems.

## 5. Conclusion

The importance of promoting the reform of spoken English teaching in our country's English teaching environment is discussed in this article. An oral teaching environment is seen to be important to facilitate the reform of spoken English teaching. However, a widespread problem in oral English instruction in colleges and universities is that the spoken conversation objects are not standardized enough, or there are no individuals with whom to converse. As a result, an intelligent oral dialogue system based on big data and neural network technology is critical, and the quality of discourse is dependent on precise oral evaluation. We began by identifying six characteristics: quality of pronunciation, fluency, content richness, issue relevance, grammar, and vocabulary richness. Second, we present a feedforward evaluation approach that connects certain TDNN layers. We can obtain richer context information and greatly reduce the number of model parameters by using feature representations of target words in different TDNN layers. Finally, we carried out a simulation test. The findings of the experiments suggest that the proposed model is accurate in evaluating spoken English and can effectively assist the reform of spoken English teaching in colleges and universities, and its performance is better than SVM by 9.2%.

## Data Availability

The data used to support the findings of this study are included within the article.

## Conflicts of Interest

The author declares that there are no conflicts of interest.

## Acknowledgments

This research was supported by the Guangdong higher education reform project "Research on the Innovation Mode of College English Micro-class under the Era of Internet Plus" (Guangdong Higher Teaching Document (2018) No. 180); Research on the Cultivation of College Students' English Autonomous Learning Ability and Strategies Under the Background of Educational Informatization (DDXK202101ZXM); Advanced English: Viewing, Listening and Speaking I—"Curriculum Ideological and Political" Demonstration Course (PPJH202118YLKC); Research on Ideological and Political Teaching Design and Practice of "Advanced English: Viewing, Listening and Speaking" Course from the Perspective of "Three- Complete Education" (ZLGC202057); and College English—Online and Offline Mixed First-class Course (PPJH202104YLKC).

## References

- [1] S. A. Asongu and N. M. Odhiambo, "Basic formal education quality, information technology, and inclusive human development in sub-Saharan Africa," *Sustainable Development*, vol. 27, no. 3, pp. 419–428, 2019.
- [2] S. F. Alfalah, "Perceptions toward adopting virtual reality as a teaching aid in information technology," *Education and Information Technologies*, vol. 23, no. 6, pp. 2633–2653, 2018.
- [3] P. Paul, A. Bhumali, and P. S. Aithal, "Indian higher education: with slant to information technology—a fundamental overview," *International Journal on Recent Researches In Science, Engineering & Technology*, vol. 5, no. 11, pp. 31–50, 2017.
- [4] V. Bošković Marković, "Traditional language teaching versus ICT oriented classroom," in *In Sinteza 2019-international scientific conference on information technology and data related research*, pp. 627–632, Singidunum university, 2019.
- [5] L. Qiu, "Computer-aided English teaching platform based on secure shell framework," *International Journal of Emerging Technologies in Learning*, vol. 14, no. 16, 2019.
- [6] U. Afini, C. Supriyanto, and R. A. Nugroho, "The development of Indonesian POS tagging system for computer-aided independent language learning," *International Journal of Emerging Technologies in Learning*, vol. 12, no. 11, 2017.
- [7] M. K. Ahmed, "Multimedia aided language teaching: an ideal pedagogy in the English language teaching of Bangladesh," *American International Journal of Social Science Research*, vol. 3, no. 1, pp. 39–47, 2018.
- [8] Y. Tong, L. Yu, S. Li, J. Liu, H. Qin, and W. Li, "Polynomial fitting algorithm based on neural network," *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 1, pp. 32–39, 2021.
- [9] J. Zhang, Y. Liu, H. Liu, and J. Wang, "Learning local-global multiple correlation filters for robust visual tracking with Kalman filter redetection," *Sensors*, vol. 21, no. 4, p. 1129, 2021.
- [10] L. Geng, "Evaluation model of college English multimedia teaching effect based on deep convolutional neural networks," *Mobile Information Systems*, vol. 2021, Article ID 1874584, 8 pages, 2021.



- [11] M. Willsey, A. P. Stephenson, C. Takahashi et al., "Puddle: a dynamic, error-correcting, full-stack microfluidics platform," in *In Proceedings of the Twenty-Fourth International Conference on Architectural Support for Programming Languages and Operating Systems*, pp. 183–197, Providence, 2019, April.
- [12] M. W. Ok, K. Rao, J. Pennington, and P. R. Ulloa, "Speech recognition technology for writing: usage patterns and perceptions of students with high incidence disabilities," *Journal of Special Education Technology*, 2020, in press.
- [13] H. Isyanto, A. S. Arifin, and M. Suryanegara, "Performance of smart personal assistant applications based on speech recognition technology using IoT-based voice commands," in *2020 International Conference on Information and Communication Technology Convergence (ICTC)*, pp. 640–645, Korea, 2020, October.
- [14] E. Armas, R. Álvarez, and G. Romero, "Aids based on speech recognition technology for people with motor disabilities and reduced mobility," *Revista Politécnica*, vol. 43, no. 1, pp. 15–22, 2019.
- [15] H. Chung, Y. K. Lee, S. J. Lee, and J. G. Park, "Spoken English fluency scoring using convolutional neural networks," in *In 2017 20th Conference of the Oriental Chapter of the International Coordinating Committee on Speech Databases and Speech I/O Systems and Assessment (O-COCOSDA)*, pp. 1–6, korea, 2017, November.
- [16] K. Evanini, M. Mulholland, E. Tsuprun, and Y. Qian, *Using an automated content scoring system for spoken CALL responses: the ETS submission for the Spoken CALL Challenge. In Proceedings of the Seventh SLaTE Workshop*, Stockholm, Sweden, 2017.
- [17] L. Hussain, I. A. Awan, W. Aziz et al., "Detecting congestive heart failure by extracting multimodal features and employing machine learning techniques," *BioMed research international*, vol. 2020, 19 pages, 2020.
- [18] W. Chu, P. S. Ho, and W. Li, "An adaptive machine learning method based on finite element analysis for ultra low-k chip package design," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, pp. 1–1, 2021, in press.
- [19] C. Yan, G. Pang, X. Bai et al., "Beyond triplet loss: person re-identification with fine-grained difference-aware pairwise loss," *IEEE Transactions on Multimedia.*, p. 1, 2021, in press.
- [20] M. Li, G. Zhou, W. Cai et al., "MRDA-MGFSNet: network based on a multi-rate dilated attention mechanism and multi-granularity feature sharer for image-based butterflies fine-grained classification," *Symmetry*, vol. 13, no. 8, p. 1351, 2021.
- [21] Y. Ding, X. Zhao, Z. Zhang, W. Cai, and N. Yang, "Multiscale graph sample and aggregate network with context-aware learning for hyperspectral image classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 4561–4572, 2021.
- [22] Z. Huang, P. Zhang, R. Liu, and D. Li, "Immature apple detection method based on improved Yolov3," *ASP Transactions on Internet of Things*, vol. 1, no. 1, pp. 9–13, 2021.
- [23] R. A. Rashid, S. B. Abdul Rahman, and K. Yunus, "Reforms in the policy of English language teaching in Malaysia," *Policy Futures in Education*, vol. 15, no. 1, pp. 100–112, 2017.
- [24] M. Y. Amin, "English language teaching methods and reforms in English curriculum in Iraq; an overview," *Journal of University of Human Development*, vol. 3, no. 3, pp. 578–583, 2017.
- [25] E. J. Hwang, B. A. Macdonald, and H. S. Ahn, "End-to-end dialogue system with multi languages for hospital receptionist robot," in *In 2019 16th International Conference on Ubiquitous Robots (UR)*, pp. 278–283, Korea, 2019, June.
- [26] O. W. Kwon, Y. K. Kim, and Y. Lee, "Task graph based task-oriented dialogue system using dialogue map for second language learning," in *Future-Proof CALL: Language Learning as Exploration and Encounters – Short Papers from EURO-CALL 2018*, pp. 153–159, Research-publishing.net, 2018.

## Research Article

# Construction of Multimedia-Assisted English Teaching Mode in Big Data Network Environment

**Hongxin Zhao** 

Normal School of Eastern Liao Ning University, Dandong 118000, China

Correspondence should be addressed to Hongxin Zhao; [zhx13470091209@126.com](mailto:zhx13470091209@126.com)

Received 27 July 2021; Revised 12 August 2021; Accepted 17 August 2021; Published 15 September 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Hongxin Zhao. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Oral English teaching is the weakest link in multimedia English teaching at this stage. English teachers are constantly exploring effective approaches to improve oral English Teaching in their own educational practice. The big data multimedia English teaching mode conforms to embark on the historical stage. Firstly, this paper constructs the big data architecture of English teaching model mining and divides the construction of the teaching model into three parts: data mining, teaching model evaluation, and improvement optimization. Data mining uses the advanced DBN network to send data into the DBN-DELM network, which significantly improves the accuracy of the multimedia assisted English construction model. The simulation results show that teaching mode construction based on big data can effectively improve students' interest in English learning; attitude; and oral English level including pronunciation, pronunciation and intonation, dialogue and communication, and oral expression and improve students' group cooperation and communication ability, autonomous learning ability, evaluation consciousness, and ability.

## 1. Introduction

The information age has become the pronoun of the 21st century, and people's learning way for English have also undergone great changes. Education, as an indispensable promoter of human development, is also making changes in line with the times [1, 2]. Especially, with the continuous integration of information technology and education, exploring the path and mode of multimedia English teaching has become an important research topic for English education experts [3]. Learning mode refers to a kind of plan or model that constitutes a learning course, selects appropriate teaching materials, and guides teaching activities in the classroom or specific environment [4]. Literature [5] defines teaching mode as teaching mode, also known as teaching structure. Teaching mode is the visualization of teaching methods, and it also is the high summary of teaching experience [6]. The definition of the concept of teaching mode given in the literature [7] is teaching mode is commonly known as the big method. It is not only a kind of teaching method but also a whole and systematic operation mode from teaching prin-

ciple, teaching content, teaching goal and task, and teaching process to teaching organization form. This operation mode is theorized. The literature concludes that teaching mode refers to the stable combination and application of two or more methods or strategies in the teaching process [8].

With the support of modern information technology and professional technology, each industry field will produce series of data. The statistics, analysis, mining, and sharing of these data will create unexpected value and wealth. Literature [9] points out that learners leave many digital footprints in online or offline learning. Through tracking and analyzing some digital footprints, we can find the rules and patterns of learners' learning behavior. Literature [10] believes that big data will bring fundamental changes to modern education, including the implementation of two-way education feedback that cannot be carried out in the past, the realization of customized teaching to realize the individual conditions, and the optimization of content, time, and methods for teaching through probability prediction. According to the literature [11], learning analysis can be used to evaluate curriculum and teaching and also to discover and grasp the

mechanism of learning process. According to the literature [12], big data technology can digitize process (including teaching materials and learning activities): the solution of teaching scores will no longer depend on the fuzzy experience of teachers, but based on the description and analysis of massive teaching mode and their solutions. Since the new curriculum standard condensed the concept of English learning activities, more and more studies have discussed the impact of the concept of English learning activities on the teaching of reading and writing, but there is less discussion on how to integrate the concept of learning activities into grammar teaching. It can be seen that the theoretical guiding significance of the concept of learning activities on other English courses other than reading, and writing still needs to be further explored. This study combines the concept of learning activities with grammar teaching, which helps to reverse the lack of practical, situational, and pan theoretical research on the theory of English learning activities, builds a teaching operation mode based on the concept of English learning activities, and improves the theory of English learning activities [13]. It is entirely possible for teachers to analyze the learning data of each student [14]. When using DBN to mine teaching big data, there are problems of low efficiency and slow convergence speed [15]. The main contributions are summarized as follows: (1) this paper uses hive spark framework to design a big data real-time data platform for English teaching to improve English teaching efficiency and change teaching mode. (2) ELM is used to optimize the DBN network, which can simplify the model structure and improve the model training efficiency. (3) This paper compares the experimental method with the flipped classroom model to verify the advantages of English teaching.

In order to change the English teaching mode, this paper constructs a multimedia English teaching mode based on DBN-DELM on cloud platform. Firstly, the big data information collection platform is constructed for collecting series of information. In the data mining layer, the advanced DBN-DELM network is used to mine effective information from history information of English teaching, and the multimedia English teaching mode is constructed. Simulations show that the efficiency and effect of English teaching are enhanced. Furthermore, this paper also puts forward some implementation suggestions to promote the application and development of big data education in China.

## 2. Structure and Design of English Teaching Platform

Based on the characteristics of big data, this paper designs the English teaching big data real-time data platform (DRTDP). Its overall architecture is as follows (see Figure 1), including four abstract and implemented technology platforms: unified data acquisition platform, unified flow processing platform, unified computing service platform, and unified data visualization platform [16]. The DRTDP platform can provide end-to-end real-time data processing capability (millisecond/minute delay), connect multiple data sources in the English teaching industry chain, extract real-time data, and provide real-time data for various English teaching big data application scenarios.

*2.1. Architecture Design of Cloud Platform.* Real-time data platform must have real-time and security. Real time performance: the real-time performance of DRTDP determines that DRTDP must complete related services in data transmission and forwarding, which is characterized by low delay and can be processed by the analysis model in time. The data delay at any stage may influence the final output of the system. Stability: DRTDP must have high stability, which have robust system in any scenario, which directly affects the usability. Security: the importance of DRTDP determines that DRTDP must have a very high security protection capability. With the increasing value of data, when an intrusion occurs, it can timely detect the specific information of the intrusion and take targeted and effective remedial measures. Therefore, the English teaching big data real-time data platform is designed as follows:

The English teaching big data real-time data platform includes information perception, wireless transmission, data mining, and analysis and visualization. Firstly, Under the guidance of big data teaching mode, on the one hand, the circulation of educational content is strengthened, which is not only the mutual reference of resources inside and outside the school but also the circulation of educational resources at home and abroad through network interaction. On the other hand, on the basis of ensuring face-to-face teaching, we should supplement and improve the traditional teaching mode, learn from each other's strengths and make up for their weaknesses, and realize the multiplication and diffusion of educational effect by means of information technology. Generally speaking, English teaching big data platform contains data layer, GPRS layer, data mining layer, and visualization layer. (1) The data layer is divided into user information, course information, practice library, video library, and forum information. According to the scope of discussion, the user information library is divided into user name, age, education, and other basic information. The forum information library is divided into chapter forum, course forum, course forum, and platform Forum. (2) The network layer is used for transmitting the sensing layer data to platform through GPRS and other communication methods. (3) The data mining layer includes algorithm library and model library. The common mining algorithms include BP neural network, SVM, and deep neural network. (4) Application layer provides information to users.

*2.2. Structure Design of Teaching Mode Based on Spark-Hive Frame.* The rapid growth of data and the urgent need related to data query make the traditional data warehouse engine difficult to conditions for storage and analysis. Hadoop as an open source architecture began to replace the traditional data warehouse, using the MapReduce programming model that can effectively segment and reasonably allocate data. Hive as an especially tool provides a query interface similar to SQL [17]. However, since the execution engine of hive compiles SQL into a series of MapReduce jobs to run, its performance cost is high. The structure of English teaching cloud platform based on Hadoop is designed as Figure 2.

Hive is convenient for developers who are familiar with MapReduce to develop mappers and reducers which work that cannot be completed by built-in mappers and reducers

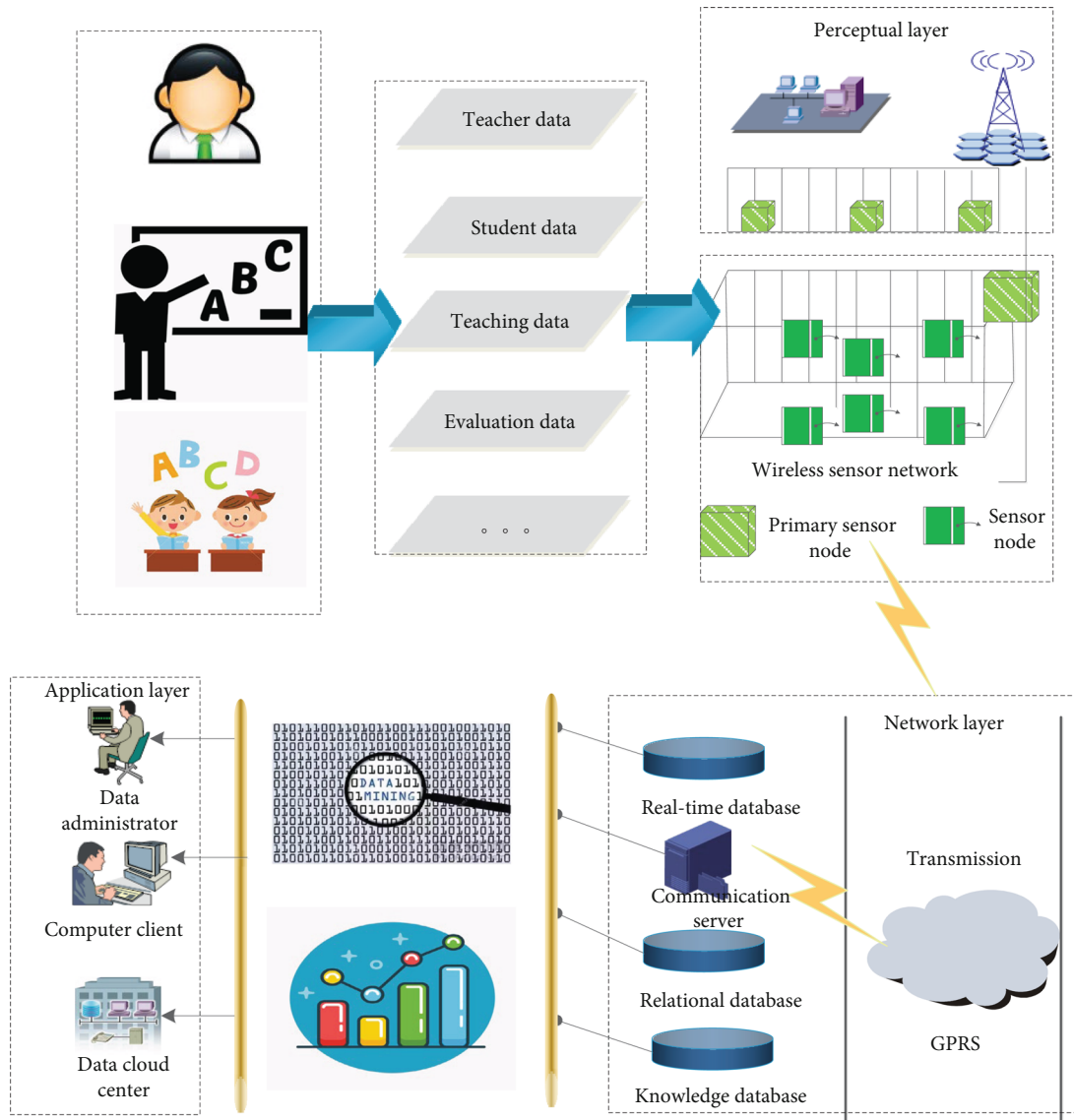


FIGURE 1: English teaching big data real-time data platform.

[18]. Hive’s table is actually HDFS’s directory/file, which divides the folder by table name. If it is a partition table, the partition value is a subfolder, which can be used directly in M/R jobs [19].

### 3. Research on Big Data Mining Algorithm Based on the DBN Neural Network

3.1. *The Process Design for Teaching Mode Construction.* To realize the intelligent construction of the multimedia English mode, this paper applies the deep learning algorithm [20–22] to the data warehouse system, uses principal component analysis [23] and normalization algorithm [24], designs a multimedia English teaching model [25], and designs an evaluation model of European media English teaching model. The deep confidence network data mining model is adjusted according to the advantages and disadvantages of the evaluation model, and its process structure is shown in Figure 3.

(1) Data sources contain two categories: internal information and external information, (1) in the teaching after class, understanding the specific situation of teachers and students, teaching content, methods, students, network English teaching and students’ participation. ② In the learning practice activities after class, we need to fully master the students’ written and oral communication ability, the ability to communicate with others in English, and the ability to read and write independently in English. ③ In English teaching system, multimedia data, and online data, we need to know something about college English learning materials, English teaching system; the latest microclass, audio, and video; English voice materials; and so on. Generally speaking, in the process of data collection and collation, it is necessary to ensure the timeliness, accuracy, and comprehensiveness of data collection to the greatest extent. These factors are the pre-conditions for the construction of ecological college English teaching mode. (2) Data preprocessing mainly processes unordered data into ordered data. (3) Data decision mainly relies

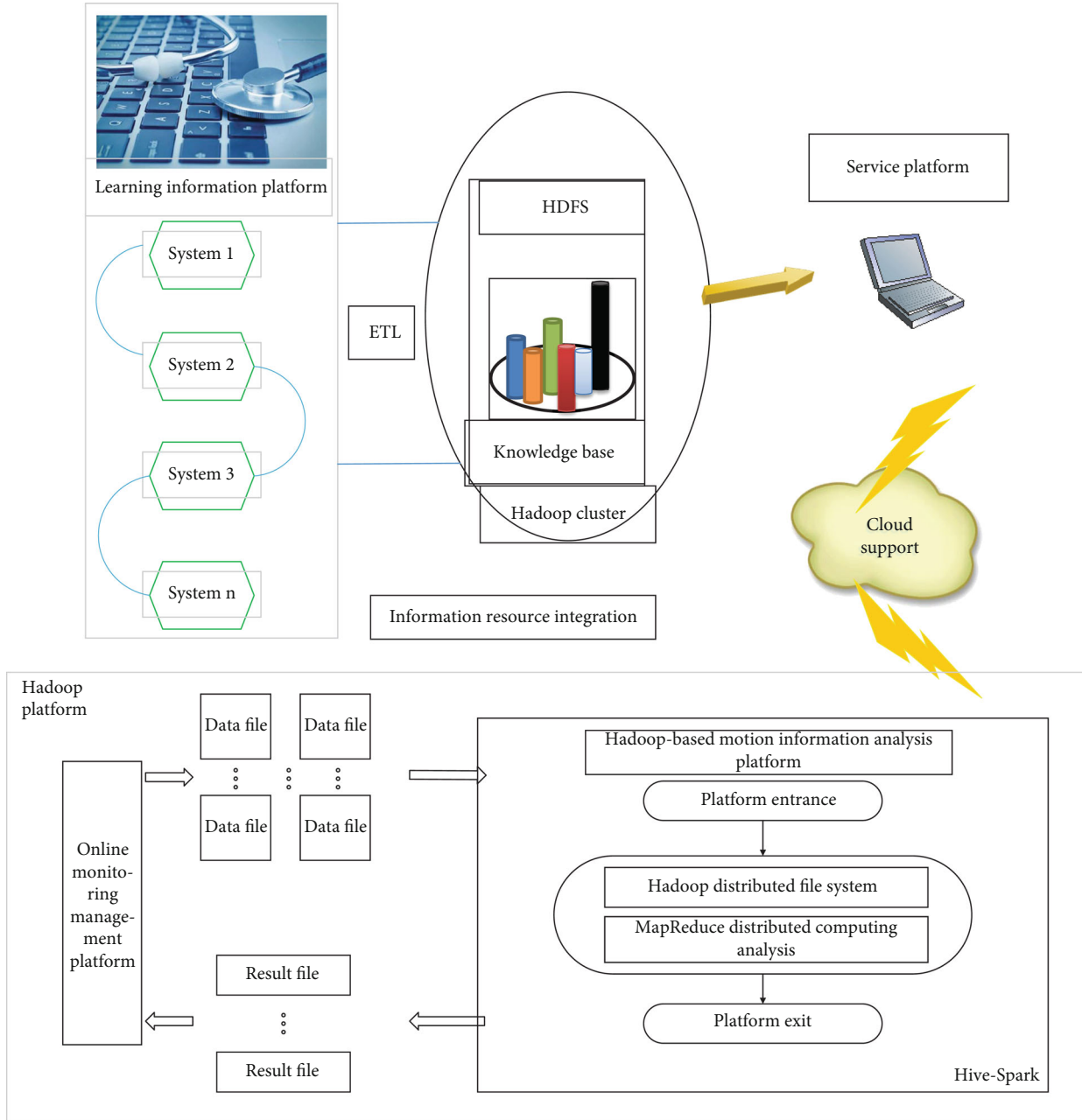


FIGURE 2: Hadoop-based architecture of operation support computing platform.

on the deep neural network. Data is directly sent to the deep neural network, and the optimal decision model can be obtained through training. (4) The model evaluation mainly evaluates the above data mining model through experiments and modifies the data decision model

3.2. *Data Mining Model Based on DBN-DELM.* Whole network: for an elm with  $n$  hidden layers, the improved dbn-delm algorithm obtains the parameters of the first  $n$  hidden layers through the DBN pretraining process and finally connects an output layer. The three-layer dbn-delm network structure is shown in Figure 4.

Deep learning itself is a branch of machine learning [26–28]. There are many similarities and differences between deep learning and traditional neural networks [29]. The similarity between the two is that deep learning adopts a hierarchical structure similar to the neural network [30–32]. The system consists of a mult-layer network composed of input layer, hidden layer (multilayer), and output layer. Only the nodes of adjacent layers are connected, and the nodes of the same layer and cross layer are not connected with each other. Each layer can be regarded as a logistic regression model. This hierarchical structure is relatively close to the structure of the human brain. There is no link between the nodes of each layer.

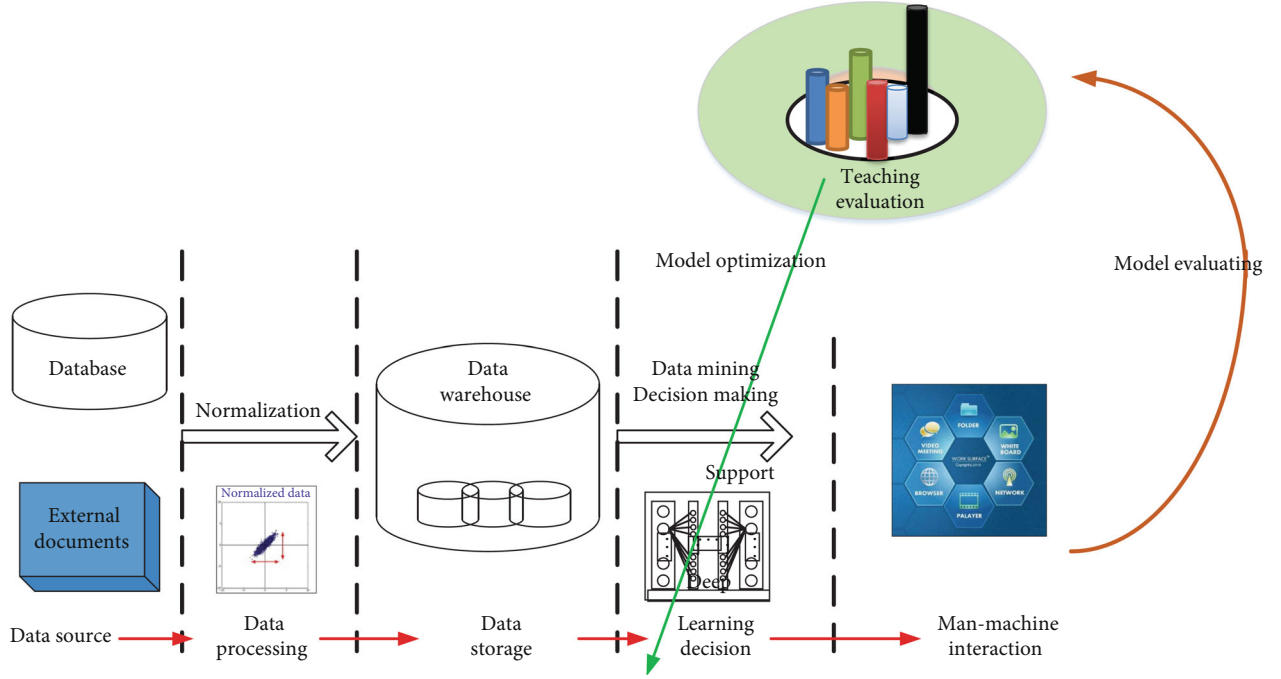


FIGURE 3: The process of teaching mode construction based on DBN-DELM.

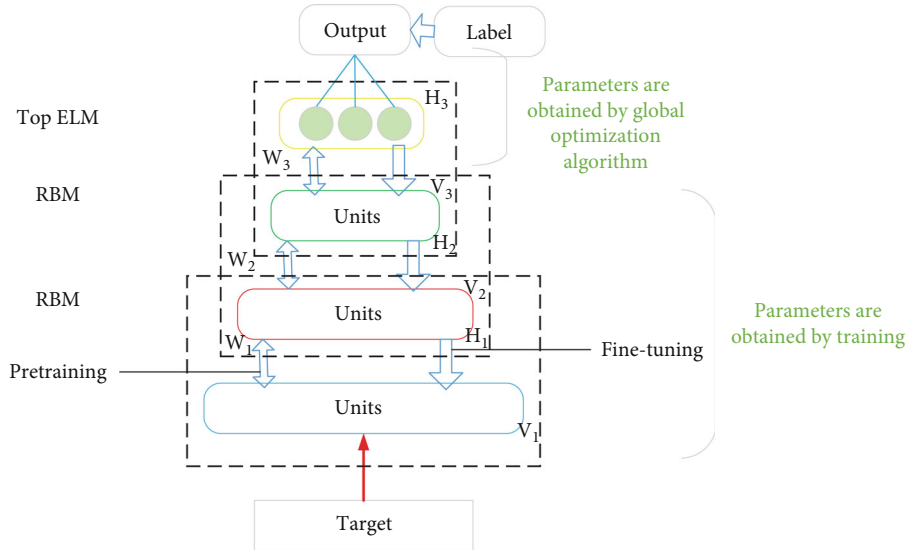


FIGURE 4: DBN network structure with the ELM algorithm.

One layer is the visual layer; that is, the input data layer ( $V$ ) and the other layer are the hidden layer ( $H$ ). If we assume that all nodes are random binary variable nodes (only 0 or 1 can be taken) and assume that the full probability distribution  $P(V, H)$  satisfies the Boltzmann distribution, we call this model restricted Boltzmann machine (RBM).

According to the pretraining method of DBN, we assume that  $v$  is the observation data,  $h$  is the feature data for hidden layer, and  $\omega$  is the link between the two layers. In the RBM unit, we can see  $m$  visible neurons and  $N$  corre-

sponding hidden layer neurons. I used the contrast divergence (CD) algorithm to learn the parameter [33] updating criteria of RBM in each layer.

$$\begin{cases} \omega_{ij} = \omega_{ij} + \varepsilon \{ \langle v_i h_j \rangle_{\text{data}} - \langle v_i h_j \rangle_{\text{recon}} \}, \\ a_i = a_i + \varepsilon \{ \langle v_i \rangle_{\text{data}} - \langle v_i \rangle_{\text{recon}} \}, \\ b_j = b_j + \varepsilon \{ \langle h_j \rangle_{\text{data}} - \langle h_j \rangle_{\text{recon}} \}. \end{cases} \quad (1)$$

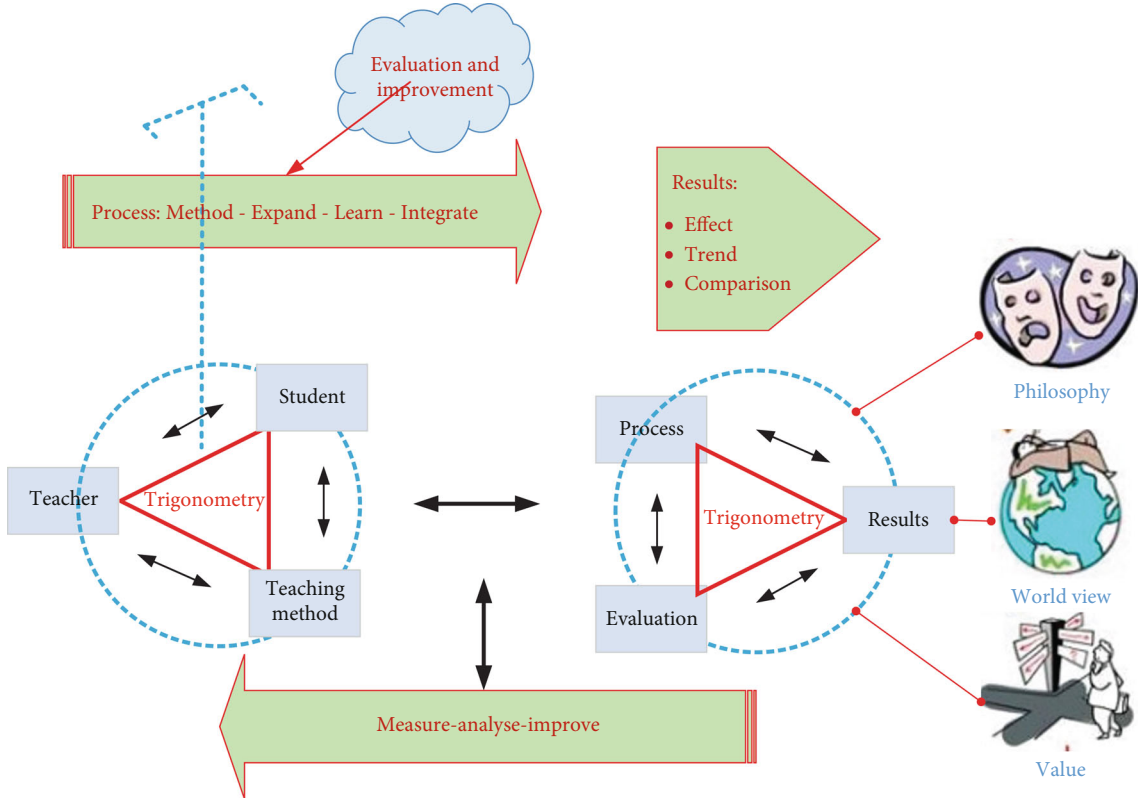


FIGURE 5: English teaching evaluation model.

TABLE 1: Cloud environment of platform.

Server type	OS	Memory	IP address	Others
Data server1	Windows 7	16GB	192.168.39.128	Opc.Tcp://10.32.42.48 : 4401
Data server2	Windows 7	16GB	192.168.39.129	Opc.Tcp://10.32.42.48 : 4401
Namenode	CentOS 7.5	64GB	192.168.39.130	Hadoop 2.6.1; spark 1.7.1
Slave node1	CentOS 7.5	64GB	192.168.39.131	Hadoop 2.6.1; spark 1.7.1
Slave node2	CentOS 7.5	64GB	192.168.39.132	Hadoop 2.6.1; spark 1.7.1

The weight vector  $\omega = \omega^1, \omega^2 \dots \omega^N$ , the offset vector  $b = b^1, b^2 \dots b^N$ , and the number of hidden layers are obtained by pretraining. It can be as follows:

$$\begin{cases} h^2 = g(\omega^1 x + b^1), \\ h^3 = g(\omega^2 h^2 + b^2), \\ \dots \\ h^{i+1} = g(\omega^i h^i + b^i). \end{cases} \quad (2)$$

Finally, the parameters from the  $n$ th hidden layer to the output layer are obtained by the ELM algorithm

$$\sum_{j=1}^N \beta_j h_j^n = o_j \quad j = 1, 2, \dots, N, \quad (3)$$

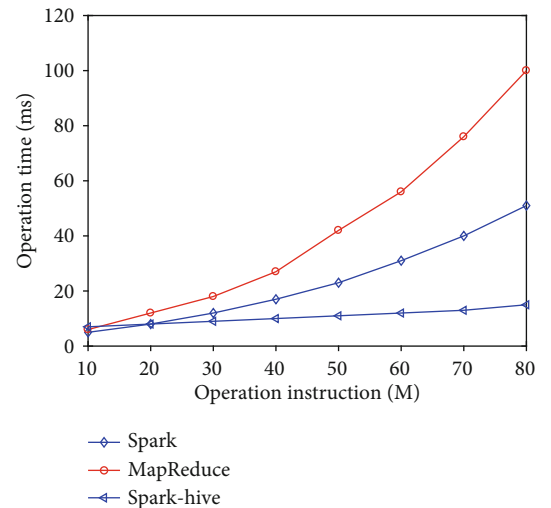


FIGURE 6: Comparison of operation efficiency of big data platforms.

TABLE 2: DBN-DELM model compared with other models.

Samples		1	2	3	4	5	6	7	8	9	10
DBN-DELM	Error	0.2	-0.1	-0.4	-0.5	0.2	0.7	0.15	0.7	-0.4	0.21
	MPE%	1.26	0.66	2.65	3.4	1.26	4.87	1.01	4.17	2.8	1.35
	RMSE	2.34%									
DBN	Error	0.4	0.6	1.3	0.2	-0.9	1.7	1.3	-0.1	2.3	0.9
	MPE%	2.56	3.97	8.33	1.36	5.66	11.2	8.78	0.6	16.1	5.81
	RMSE	6.44%									
BPNN	Error	1.3	0.55	1.48	3.25	-3	1.53	1.6	4.6	-2.6	3.4
	MPE%	9.09	3.78	10.6	28.3	15.9	11.2	12.1	37.7	15.4	28.1
	RMSE	17.22%									

where  $\beta_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{iN}]$  is the parameter vector between the  $n$ th hidden layer neurons and the output layer neurons, and  $N$  is the number of  $n$ th hidden layer neurons.

The difference between the DBN-DELM algorithm and IDBN algorithm is that IDBN simply stacks and combines the DBN and elm algorithm, while the DBN-DELM algorithm gets better initialization parameters through the unsupervised process of DBN through the random initialization of the deep elm algorithm and then gets the top parameters through the global optimization of the elm algorithm, which speeds up the training speed of DBN, and the accuracy of the elm algorithm is improved.

**3.3. Teaching Mode Evaluation Model.** The realization of multimedia teaching effect is inseparable from the production of high-quality multimedia courseware. Without the design and production of multimedia courseware, it is impossible to produce the expected ideal classroom teaching effect. The realization of multimedia English teaching mode effect is inseparable from the appropriate multimedia teaching methods. To correctly evaluate the advantages and disadvantages of English teaching mode based on DBN-DELM, the teaching mode evaluation model is as shown in Figure 5.

The selection and determination of evaluation content are the key to the implementation of teaching evaluation. According to the new English curriculum standard, teachers should consider comprehensively and analyse the actual situation of students, choose the evaluation content in an appropriate way, and devote oneself to the diversification of the evaluation content, so as to avoid overemphasizing the examination of language knowledge and ignoring the practical application of language.

Referring to the new English curriculum standard, the author divides the evaluation content of English teaching mode based on DBN-DELM into two parts: the oral English ability and the evaluation of learning behavior. Language competence is evaluated from two aspects. The MAE index is as selected:

$$\min_{\theta \in \Omega} \text{MAE} = \frac{1}{M} \sum_{i=1}^M |T_e^i(iT | \theta) - T_e^i(iT)|, \quad (4)$$

where  $i$  is the length of the  $i$ th time step.  $M$  is the total number

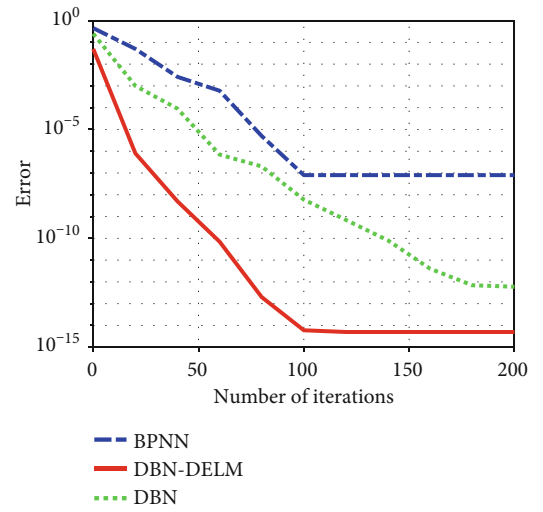


FIGURE 7: The error convergence comparison of three models.

of the sample time that are used to train or validation.  $\Omega$  is the solution space of  $\theta$ .

## 4. Simulation and Results Analysis

**4.1. Simulation Environment of Cloud Platform.** In this experimental environment, spark is used to compare with traditional MapReduce to test teaching mode in the professional model. Hadoop 2.6 big data platform is adopted, and yard is used to manage the cluster with two hosts. Configuration of each host: each machine has 4 memory modules, each 4 GB, a total of 32 GB of memory, the amount of data is 2 GB, and a total of 1 million pieces of data. To improve the efficiency of test operation, Hadoop cluster consists of one Namenode (host name is node1) node and four datanodes (host name is node 1, node 2, node 3, node 4). Hive runs on node 1 node, and zookeeper runs on node 2, node 3, and node 4 node with odd number of nodes. The configuration of each node is shown in Table 1.

In order to improve the convergence speed of DBN, this paper mainly optimizes and improves the deep confidence network DBN and applies it to the multimedia English teaching system. Therefore, this paper has carried out three experiments:



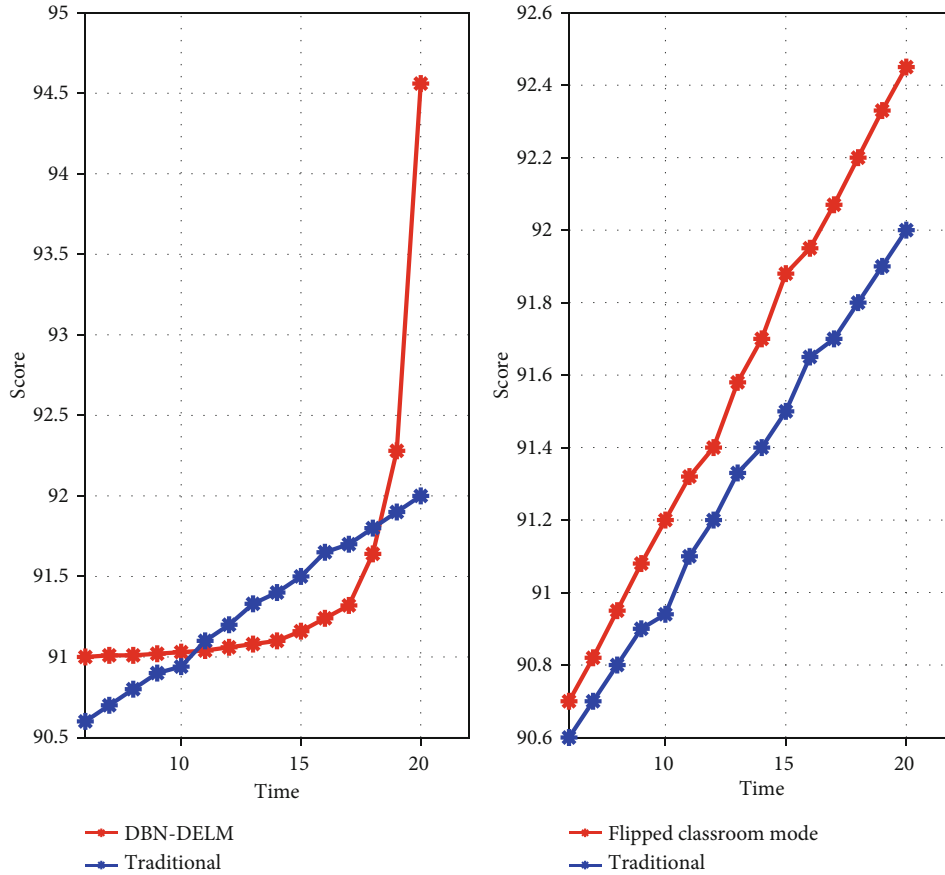


FIGURE 8: The score results compared with flipped classroom teaching mode.

(1) apply hive spark framework to the multimedia big data English teaching system, optimize the system, and improve the mining efficiency of the system. At this time, compare this framework with spark and MapReduce framework, (2) the elm algorithm is used to optimize DBN. At this time, DBN-DELM is compared with DBN and BPNN to verify the efficiency of the algorithm. (3) The multimedia English teaching system is compared with the traditional flipped classroom to verify the advantages of the system.

**4.2. The Evaluation of Hadoop Platform.** The efficiency of big data platform is not verified. This paper selects the common Hadoop framework platform for comparison. The frameworks of Hadoop are shown in Figure 6.

From the information shown in Figure 6, the time required for hive is basically proportional to the amount of data. The call of MapReduce by hive through SQL statements will lead to network and read-write disk overhead, which affects the query efficiency. For the same amount of data, hive is more efficient. In addition, for hive, the query time in spark mode is one tenth of hive-stand-alone and spark hive. In the client mode, there is not much time difference. With the increase of the number of data, hive is more stable. The advantage of spark framework query ability will be more and more obvious. Overall, the combination of spark hive is better than hive.

**4.3. The Superiority Validation of Teaching Mode Based on DBN-DELM.** To verify the superiority of the teaching mode, two groups of experiments are designed: (1) compared with DBN [34] and BPNN [35], the accuracy of the model is verified. (2) Compared with the Hadoop platform of MapReduce, the efficiency of big data platform is verified.

From the prediction results, we can see that the construction of teaching mode mined by the DBN-DELM algorithm has higher prediction accuracy than traditional DBN and BPNN, and the comparison results are shown in Table 2. The accuracy of the prediction results is evaluated by calculating the mean square deviation of the prediction results, that is, the average relative error (PMSE) value in the graph. Intuitively, the error of the improved algorithm is the smallest, between 1.5% and 2.5%, and the prediction effect is better. Secondly, due to the introduction of interest degree, it can effectively remove the wrong strong association rules, the prediction effect is second, and the average relative error is maintained between 5% and 10%. Compared with the former two optimization algorithms, the traditional BPNN algorithm has the worst prediction effect, with the average relative error between 10% and 20%.

To further verify the convergence index for DBN-DELM models, the MAE index convergence of the three models is shown in Figure 7. The DBN-DELM algorithm can converge faster, and the final convergence value is much smaller than

DBN and BPNN algorithms; The DBN-DELM mining model is further verified by the comparison figure.

To verify the superiority of the DBN-DELM-based model, I randomly select three classes for comparison. One of the classes is the traditional teaching mode, as the control class. The others as the experimental group, respectively, applied the DBN-DELM-based model and flipped classroom mode. The comparison data are shown in Figure 8. The 20 English tests were selected as the control. The results show that the scores of the traditional control group are rising steadily with the recommendation of time, and the multimedia English teaching mode based on flipped classroom can also achieve good teaching results over time. Flipped classroom reflects students' awareness of autonomous learning to a certain extent. Under the traditional grammar teaching mode, students are used to accepting and executing the learning tasks assigned by teachers and rarely actively reflect and evaluate their own grammar learning methods and effects. However, this experimental teaching cycle is short, and it is unable to significantly improve students' internal awareness of autonomous learning grammar in a short time; this may be one of the reasons why the experimental class students' metacognitive strategies have not improved significantly. However, the effect of multimedia English teaching mode based on DBN-DELM is not obvious in the early stage. After 15 examinations, the score will rise exponentially, which has a good effect. Furthermore, the English teaching model based on DBN-DELM plays a significant role in improving students' grammar ability and the use of grammar learning strategies, especially in improving the use of learners' emotional strategies, communication strategies, and cognitive strategies.

## 5. Conclusion

In the past, the teaching model used in college English teaching pays too much attention to students' classroom performance, but ignores students' practical ability of English language, and does not create practical communication opportunities for students, which leads to the disconnection between learning and practice. Therefore, this paper designs the construction of the DBN-DELM-based model. Through the design of big data platform architecture and hive technology application, the efficiency and speed of big data mining increase. The combination of elm and DBN can improve the speed and accuracy of model training convergence. In addition, the multimedia English teaching mode based on big data attaches importance to the relevance and openness of its constituent elements and implements the practical application awareness into the whole teaching process. It uses big data technology to reasonably associate the elements involved in teaching and each teaching environment, so as to ensure the effective sharing of information. Finally, the accuracy of the model and solve the problem of students' scores in Multimedia English teaching environment is verified. In addition, using the deep loop network to improve multimedia English teaching methods is the focus of the next research.

## Data Availability

The data used to support the findings of this study are included within the article.

## Conflicts of Interest

The author does not have any possible conflicts of interest.

## Acknowledgments

This study was supported without any funding.

## References

- [1] W. Li and X. Fan, "Construction of network multimedia teaching platform system of college sports," *Mathematical Problems in Engineering*, vol. 2021, no. 8, Article ID 6304703, 11 pages, 2021.
- [2] L. Diao and P. Hu, "Deep learning and multimodal target recognition of complex and ambiguous words in automated English learning system," *Journal of Intelligent & Fuzzy Systems*, vol. 40, no. 4, pp. 7147–7158, 2021.
- [3] Z. Liu, L. Li, X. Fang et al., "Hard-rock tunnel lithology prediction with TBM construction big data using a global-attention-mechanism-based LSTM network," *Automation in Construction*, vol. 125, no. 3, article 103647, 2021.
- [4] D. Wang, J. Fan, H. Fu, and B. Zhang, "Research on optimization of big data construction engineering quality management based on RNN-LSTM," *Complexity*, vol. 2018, Article ID 9691868, 16 pages, 2018.
- [5] X. Yan, A. Sedykh, W. Wang, B. Yan, and H. Zhu, "Construction of a web-based nanomaterial database by big data curation and modeling friendly nanostructure annotations," *Nature Communications*, vol. 11, no. 1, 2020.
- [6] Y. H. Kim, D. C. Jeong, K. Pak et al., "Gene network inherent in genomic big data improves the accuracy of prognostic prediction for cancer patients," *Oncotarget*, vol. 8, no. 44, pp. 77515–77526, 2017.
- [7] D. Chen, "Research on traffic flow prediction in the big data environment based on the improved RBF neural network," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 4, pp. 2000–2008, 2017.
- [8] W. J. Shyr, L. W. Zeng, C. K. Lin, C.-M. Lin, and W.-Y. Hsieh, "Application of an energy management system via the internet of things on a university campus," *EURASIA Journal of Mathematics, Science and Technology Education*, vol. 14, no. 5, 2018.
- [9] B. Tran, P. Straka, M. O. Falster, K. A. Douglas, T. Britz, and L. R. Jorm, "Overcoming the data drought: exploring general practice in Australia by network analysis of big data," *Medical Journal of Australia*, vol. 209, no. 2, pp. 68–73, 2018.
- [10] K. Dehghanpour and H. Nehrir, "Real-time multiobjective microgrid power management using distributed optimization in an agent-based bargaining framework," *IEEE Transactions on Smart Grid*, vol. 9, no. 6, pp. 6318–6327, 2018.
- [11] C. Arrighi, L. Rossi, E. Trasforini et al., "Quantification of flood risk mitigation benefits: A building-scale damage assessment through the RASOR platform," *Journal of Environmental Management*, vol. 207, pp. 92–104, 2018.
- [12] C. Ho, J. Ali, and K. Caals, "Ensuring trustworthy use of artificial intelligence and big data analytics in health insurance,"

- Bulletin of the World Health Organization*, vol. 98, no. 4, pp. 263–269, 2020.
- [13] M. Zhang, “Application of brain neuroscience in the discussion of multimedia English teaching mode,” *NeuroQuantology*, vol. 16, no. 5, 2018.
- [14] M. Amoasi Acquah, D. Kodaira, and S. Han, “Real-time demand side management algorithm using stochastic optimization,” *Energies*, vol. 11, no. 5, p. 1166, 2018.
- [15] C. Wang, Y. Xia, D. Wang, Z. Niu, Y. Liu, and C. Yu, “Dynamic risk assessment of deep-water dual gradient drilling with SMD system using an uncertain DBN-based comprehensive method,” *Ocean Engineering*, vol. 226, article 108701, 2021.
- [16] Y. Bo and Y. Meifang, “Construction of the knowledge service model of a port supply chain enterprise in a big data environment,” *Neural Computing and Applications*, vol. 33, no. 5, pp. 1699–1710, 2021.
- [17] Y. Deng, W. Jiang, Z. Tang, Z. Ling, and Z. Wu, “Long-term changes of open-surface water bodies in the Yangtze River basin based on the Google earth engine cloud platform,” *Remote Sensing*, vol. 11, no. 19, p. 2213, 2019.
- [18] C. Xu, J. Yang, K. Yin, and H. Yu, “Optimal construction of virtual networks for cloud-based MapReduce workflows,” *Computer Networks*, vol. 112, pp. 194–207, 2017.
- [19] X. Chen and N. Metawa, “Enterprise financial management information system based on cloud computing in big data environment,” *Journal of Intelligent & Fuzzy Systems*, vol. 39, no. 4, pp. 5223–5232, 2020.
- [20] K. Hao, “Multimedia English teaching analysis based on deep learning speech enhancement algorithm and robust expression positioning,” *Journal of Intelligent & Fuzzy Systems*, vol. 39, no. 2, pp. 1779–1791, 2020.
- [21] S. Qi, X. Ning, G. Yang et al., “Review of multi-view 3D object recognition methods based on deep learning,” *Displays*, vol. 69, article 102053, 2021.
- [22] W. Chu, P. S. Ho, and W. Li, “An adaptive machine learning method based on finite element analysis for ultra low-k chip package design,” *IEEE Transactions on Components, Packaging and Manufacturing Technology*, p. 1, 2021, In press.
- [23] H. Abdi and L. J. Williams, “Principal component analysis,” *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 4, pp. 433–459, 2010.
- [24] Z. Wang, P. Zhang, W. Sun, and D. Li, “Application of data dimension reduction method in high-dimensional data based on single-cell 3D genomic contact data,” *ASP Transactions on Computers*, vol. 1, no. 2, pp. 1–6, 2021.
- [25] L. Liang, Q. Yin, and C. Shi, “Exploring proper names online and its application in English teaching in university,” *ASP Transactions on Computers*, vol. 1, no. 1, pp. 24–29, 2021.
- [26] M. Zhao, A. Jha, Q. Liu et al., “Faster mean-shift: GPU-accelerated clustering for cosine embedding-based cell segmentation and tracking,” *Medical Image Analysis*, vol. 71, article 102048, 2021.
- [27] W. Chu, *Studies on the effects of wiring density on chip package interaction and design optimization with machine learning*, [Ph.D. thesis], The University of Texas at Austin, 2021.
- [28] J. Zhang, J. Sun, J. Wang, and X. G. Yue, “Visual object tracking based on residual network and cascaded correlation filters,” *Journal of ambient intelligence and humanized computing*, vol. 12, no. 8, pp. 8427–8440, 2021.
- [29] L. Huang, G. Xie, W. Zhao, Y. Gu, and Y. Huang, “Regional logistics demand forecasting: a BP neural network approach,” *Complex & Intelligent Systems*, 2021, In press.
- [30] J. Zhang, W. Wang, C. Lu, J. Wang, and A. K. Sangaiah, “Light-weight deep network for traffic sign classification,” *Annals of Telecommunications*, vol. 75, no. 7-8, pp. 369–379, 2020.
- [31] M. Zhao, Q. Liu, A. Jha et al., “VoxelEmbed: 3D instance segmentation and tracking with voxel embedding based deep learning,” *arXiv preprint arXiv: 2106.11480*, 2021, In press.
- [32] Q. Liu, I. M. Gaeta, M. Zhao et al., “ASIST: ANNOTATION-FREE synthetic instance segmentation and tracking by adversarial simulations,” *Computers in Biology and Medicine*, vol. 134, article 104501, 2021.
- [33] M. Miao, W. Cai, and X. Li, “Parameter estimation of gamma-gamma fading with generalized pointing errors in FSO systems,” *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 1301878, 14 pages, 2021.
- [34] O. Taylan, M. R. Kabli, C. Porcel, and E. Herrera-Viedma, “Contractor selection for construction projects using consensus tools and big data,” *International Journal of Fuzzy Systems*, vol. 20, no. 4, pp. 1267–1281, 2018.
- [35] W. Huang, “Research on the revolution of multidimensional learning space in the big data environment,” *Complexity*, vol. 2021, no. 4, Article ID 6583491, 12 pages, 2021.

## Research Article

# Stock Trend Prediction Algorithm Based on Deep Recurrent Neural Network

Ruochen Lu <sup>1</sup> and Muchao Lu<sup>2</sup>

<sup>1</sup>The University of Queensland, Brisbane 4075, Australia

<sup>2</sup>Taiyuan University of Technology, Taiyuan 030024, China

Correspondence should be addressed to Ruochen Lu; [ruochen.lu@uqconnect.edu.au](mailto:ruochen.lu@uqconnect.edu.au)

Received 2 August 2021; Revised 23 August 2021; Accepted 24 August 2021; Published 14 September 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Ruochen Lu and Muchao Lu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the return of deep learning methods to the public eye, more and more scholars and industry researchers have tried to start exploring the possibility of neural networks to solve the problem, and some progress has been made. However, although neural networks have powerful function fitting ability, they are often criticized for their lack of explanatory power. Due to the large number of parameters and complex structure of neural network models, academics are unable to explain the predictive logic of most neural networks, test the significance of model parameters, and summarize the laws that humans can understand and use. Inspired by the technical analysis theory in the field of stock investment, this paper selects neural network models with different characteristics and extracts effective feature combinations from short-term stock price fluctuation data. In addition, on the basis of ensuring that the prediction effect of the model is not lower than that of the mainstream models, this paper uses the attention mechanism to further explore the predictive *K*-line patterns, which summarizes usable judgment experience for human researchers on the one hand and explains the prediction logic of the hybrid neural network on the other. Experiments show that the classification effect is better using this model, and the investor sentiment is obtained more accurately, and the accuracy rate can reach 85%, which lays the foundation for the establishment of the whole stock trend prediction model. In terms of explaining the prediction logic of the model, it is experimentally demonstrated that the *K*-line patterns mined using the attention mechanism have more significant predictive power than the general *K*-line patterns, and this result explains the prediction basis of the hybrid neural network.

## 1. Introduction

For investors, stocks meet different investment desires and investment needs, expand the range of investment options, expand investment channels, to some extent meet the possibility of investors to obtain the corresponding income, and to some extent enhance the flexibility and liquidity of capital [1]. If we look at the enterprise side, the stock can play an important role in the management and development of the joint-stock enterprise, which is conducive to the establishment and improvement of the self-restraint and self-development of the enterprise management mechanism [2]. For the country, stocks are also a great tool to counteract inflation. Stocks have three main characteristics: (1) nonre-

turnability: once sold, stocks cannot be returned to the company and cannot be requested to be refunded but can only be sold to a third party through the secondary market; (2) uncertainty of earnings: the profit and loss of stocks depend on the company's operation and the stock exchange market, both of which are uncertain and changeable, so investors need to take greater risks; and (3) speculative: the stock market fluctuates frequently and the market price is unstable. Market prices are unstable and speculative, so stocks are very risky [3]. There are many reasons that affect the volatility of stock prices, and these frequently changing factors cause stock market volatility [4]. The objective risks of the stock market can bring gains to investors and at the same time can cause economic losses and may also

negatively affect the operating conditions of shareholding companies and even bring side effects to the national economic construction.

These problems are inevitable, so stock trend prediction has become an issue of great concern for all parties [5]. The research of stock trend prediction has also become an applied research direction of financial big data, and many scholars have adopted deep neural network methods to predict the stock trend, which has become one of the popular research problems in the current academic field. In recent years, with the rapid development of computer technology, the deep neural network [6–8] has become the key research object nowadays, and the application field has been expanded and extended, including the financial information field [9]. At present, the financial market occupies a pivotal position in the whole economic system of the country, and stocks are an important component of the financial market, so buying stocks has become a popular financial management method nowadays, and a stock analysis software with various performance can meet the real needs of investors [10]. This paper designs and implements a stock trend prediction system by combining historical stock trading data, news data, and investor sentiment to recommend stocks and stock trend prediction services for investors to reduce or avoid investment risks, thus bringing investors relatively stable economic returns. Due to the continuous innovation of machine learning techniques [11–13], more and more researchers are turning to the use of machine learning techniques to analyze stock data and create well-performing methodological models to predict stock movements in the future. Stock market prediction models are built by learning from historical price data to predict future prices [14]. Common machine learning algorithms such as logistic regression, genetic algorithms, and support vector machines have been used with good results in forecasting [15]. With the rise of neural network technology, building deep neural networks to portray stock prices and predict stock movements has received a lot of attention, and some scholars have conducted in-depth research in this area [16]. In order to improve the accuracy of stock trend prediction, many improved algorithms and optimization strategies have appeared one after another and have been successfully applied in practice.

Compared with existing stock trend prediction models, the improved model used in this paper combines stock price trend, news information, and investor sentiment for prediction, not only using trading data in the stock market but also taking into account the influence of financial and political news and stock forum speech on the stock market. In this work, a framework for predicting stock price trends using financial news and sentiment dictionaries combines a two-stream gated recurrent unit for stock price trend prediction and a Stock2Vec embedding model trained on stock news and sentiment dictionaries. Firstly, we propose a plain Bayesian model based on the sentiment classification of stock forum speech and experimentally confirm that the classification effect is better when using the plain Bayesian classifier, which can obtain the investor sentiment more accurately and lay the foundation for the establishment of

the whole stock trend prediction model. Secondly, on top of the original LSTM, this model is constructed by a mixture of Bi LSTM [17] and CLSTM, with Bi LSTM extracting stock trading data and investor sentiment index-related features and CLSTM integrating and processing the contextual features of the news, and finally outputting the prediction results through the fully connected layer. In the experimental model, the stock trend is experimented with using the classification method [18–20], and the classification is obtained as the probability of the stock going up and the probability of the stock going down. The experiments use the CSI 300 stock data as the data set. The prediction effect is evaluated by accuracy and return, and the experimental results show that compared with a single LSTM prediction model, the proposed method has a certain improvement in the accuracy of stock trend prediction and can predict stock trend accurately and effectively to a certain extent. At the same time, a deep neural network stock trend prediction [21] system is designed and implemented, and the trained prediction model is uploaded to the stock prediction module. By analyzing the requirements of the stock trend prediction system and designing each functional module, the whole stock trend prediction system is completed after development and testing. Investors can make stock selection and investment with the help of this system.

## 2. Related Work

The researchers constructed a denoising hybrid stock price prediction model based on a decision tree, which first extracts relevant features from stock data, then uses the principal component analysis algorithm to dimensionally reduce the features with the decision tree algorithm, and the dimensionally reduced data [22] is fed into the fuzzy model to predict stock prices [23]. The researchers created a Bayesian neural network model that does not require preprocessing operations and cycle analysis of the data but simply feeds market prices and technical indicators into the prediction model, which is used to predict the future closing price of the stock [24]. The researchers constructed a support vector machine model with a genetic algorithm to optimize the data by dimensionality reduction and used a quantitative stock selection method to empirically analyze its stock selection performance and prediction accuracy in the short and medium to long term, respectively [25]. The researchers first used wavelet decomposition of stock price series to screen out the low- and high-frequency information in the nonstationary time series and then constructed an ARIMA model from the high-frequency information and fitted the SVM model to the low-frequency information to obtain better results [26].

The researchers performed feature construction based on relevant technical indicators and used data mining techniques for feature modelling [27]. The main method relies on the idea of maximizing returns and proposes a support vector machine for genetic parameter search with AUC values under the ROC curve, which solves the problem of poor availability of forecasting with traditional methods. The researchers used Tensor Flow, a Google AI learning

system, and thus built a multiperceptron MLP neural network model for predicting each closing price, comparing Tensor Flow with a traditional BP neural network [28] based on the stock price prediction problem [29]. The researchers used a different number of features and multivariate data preprocessing to study the effect of long- and short-term memory networks (LSTM networks) on the prediction results, and the LSTM models increased the accuracy of stock return prediction compared to random prediction methods [30]. Researchers use LSTM to predict stock prices and propose variable step integration methods and an improved MSE loss function with improved prediction performance, but the drawback is that no generalized optimal step range is derived.

Many factors can affect stock markets and cause market volatility, such as global economic conditions, domestic macroeconomic factors, and highly correlated foreign stock markets. However, most time series models use stock indices as the only factor for prediction and do not consider more variables, while the opposite can lead to better prediction results [31]. Therefore, some scholars have developed sentiment analysis models for predicting the correlation between investor sentiment and financial markets, and researchers have studied the relationship between public sentiment states obtained from Twitter and the Dow Jones Industrial Average (DJIA) [32]. Two sentiment analysis models were used to analyze the text content of daily Twitter feeds to obtain and analyze changes in the public's sentiment. Researchers' sentiments were classified as positive and negative, while GPOMS meticulously classified sentiments into six categories, including calm, alert, sure, vital, kind, and happy, with six different dimensions to measure sentiments [33]. Through the Granger causality test, a close relationship between public sentiment and the Dow Jones Average Index (DJIA) was found. Next, using a self-organizing fuzzy neural network model, the public sentiment time series was used as the independent variable, and the model was found to be effective in predicting the change of DJIA closing price, which can largely improve the accuracy of DJIA prediction [34].

The researchers propose a framework for predicting stock price trends using financial news and sentiment dictionaries, combining a two-stream gated recurrent unit (TGRU) for stock price trend prediction and a Stock2Vec embedding model trained on stock news and sentiment dictionaries [35]. The researchers predict stock trends through the impact of financial news on stock market sentiment, using more expressive features to represent the text and augmenting existing text mining methods by including market feedback as part of the feature selection process. Powerful feature selection can greatly improve classification accuracy when used in conjunction with complex feature types, and the approach allows the selection of semantically relevant features, thus reducing the problem of overfitting when applying machine learning methods. It can also be transferred to any other application area that provides textual information and corresponding effect data. In summary, many applications have been made by many scholars using neural network models for stock prediction analysis. But most of them have been analyzed for foreign stock markets,

but there is still very little research on the domestic stock market, which may be related to the fact that the domestic stock market was established late and many aspects are still not well developed. However, the Chinese stock market has grown to have more than 3,000 listed companies with a circulating market capitalization of tens of trillions of yuan, which occupies a significant share in the whole national financial system, so it is significant to study the domestic stock market. In this paper, we construct a long- and short-term memory neural network model to predict the domestic stock trend.

## 2.1. Deep Recursive-Based Stock Trend Prediction Model

*2.1.1. Factors Influencing Stock Price and Fundamental Analysis Method.* Like other countries in the world, the Chinese stock market is complex and dynamic. From the specific business environment of listed companies, the transparency of financial statements, and the psychological sentiment of investors to the national policies and regulations, unexpected news events, and the world economic situation, all of them have a certain impact on it. Because the stock market is affected by many factors, it is difficult to predict and determine the future trend of stocks, and there is a large risk and unknown. At the same time, a stock is only marketable security, which does not have real value itself. Stock price refers to the trading price of a stock in the securities market, and dividend income is obtained by buying and selling stocks. Stock prices are influenced by various economic and political factors and other external circumstances. Technical analysis is a basic forecasting method that determines the future movement of the stock market by analyzing and judging graphical charts and related technical indicators.

It is a method of making judgments about stock price changes and stock trends based on market behavior itself, combined with theories and methods related to psychology and statistics, and based on historical stock trading data such as existing prices, rates of change, and volume in the stock market, as well as combining investors' subjective judgments and analysis to find patterns. The actual trading history of a particular stock or "average" is usually recorded in graphical form, and then possible future trends are inferred from that history. Technical analysis is based on the theory that market behavior contains all information and is based on three theories: that all information is fair and open, that stock prices move along a trend, and that historical stock prices repeat themselves. It mainly includes the indicator,  $K$ -line, pattern, and wave methods. Because the stock system is highly nonlinear and stochastic, it requires the investor to analyze graphical movements and data tables to obtain forecasts, while understanding the role of relevant parameters and corrections. Therefore, this method is not applicable in today's increasingly large and complex stock market. In addition to low efficiency, difficulty, and overreliance on manual experience, it also has a series of problems such as poor integrity of stock content information, redundancy of feature data, low utilization of stock data, poor results, and poor generalization, which brings certain difficulties in predicting stock prices and makes it difficult to meet the needs

of market development. Fundamental analysis is a method to calculate the intrinsic value of a stock by looking at the basic economic factors that may affect the stock price, i.e., fundamentals.

Relevant factors considered include turnover, revenues and expenses, financial statements, the company's growth prospects, competitive factors facing the company, and the expected return on equity or assets for the industry. The purpose of this analysis is to determine a value for the stock that takes all of these potential factors into account. Investors rely on a variety of financial tools to determine a company's past, present, and future profitability in order to make investment decisions. Some of the key financial statements that investors rely on include balance sheets, cash flow statements, and income statements. Accurate financial statements are used to analyze the corresponding value of a company. In order to determine the accuracy of such statements, investors rely on independent auditors' reports to ensure the accuracy of the financial statements. The method is considered a long-term investment approach because it does not take into account short-term pricing and trading fluctuations. The method also involves some element of forward-looking expectations, and it may take some time to realize the intrinsic value, so forward-looking information needs to be evaluated before considering the use of this method. Meanwhile, the method requires a comprehensive analysis of both macro- and microinformation. Micro is difficult to guarantee access to real and valid financial information because of the asymmetry of corporate information, as shown in Figure 1. The specific parameters can be from the raw data collected by sensors, which are continuous data sequences generated by user movements over a period of time. Macroscopically, it is also difficult to predict the national plan support policies and key development industries. Therefore, the analysis method still needs to be improved in terms of accuracy and timeliness, and it relies too much on the ability of analysts, and the analysis method is relatively difficult to apply and has certain limitations, especially for ordinary stock market investors, with poor feasibility and universality. To sum up, the fundamental analysis method mainly analyzes the long-term trend of stocks, which is difficult to obtain, organize, categorize, and analyze information, requires high analytical ability of investors, and does not allow reliable and accurate short-term forecasting.

**2.1.2. Stock Trend Forecasting Model.** The operational status of a listed company can be reflected by news information about the company's stock. Based on the common feature that news information is highly correlated with stock movements, people often rely on news information to forecast stock movements. Due to the time-sensitive nature of news, the information in news articles has a short time effect on the stock market, and in general, recent news has a large impact on stock movements. In order to encode stock-related news information, data preprocessing work is first performed on the text to reduce the repetition rate and correct the comment data with wrong formatting to improve the integrity and quality of the comment data. Unreasonable cases such as noisy data, such as null values and special sym-

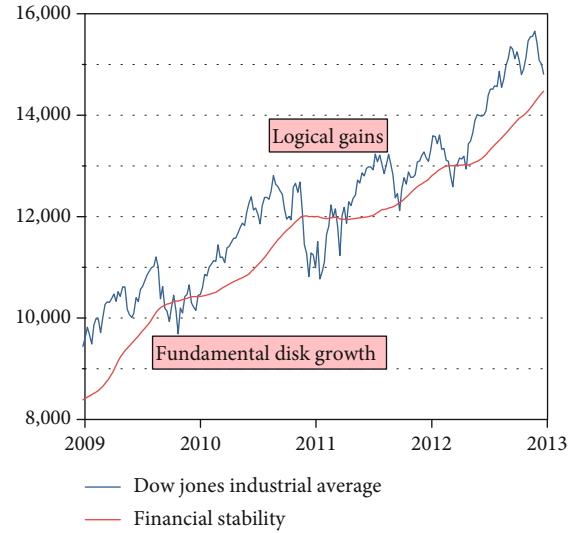


FIGURE 1: Correlation between stock price and financial factors.

bols, need to be deleted and processed. The authenticity and reliability of the research results are ensured by data preprocessing operations.

- (1) *Handling Missing Data.* The initial data obtained may suffer from data loss, which may be due to incorrectly unentered data, data that is inconsistent with data from other records and therefore deleted, data that was ignored at the time of entry, and unrecorded data changes. Missing data can be handled in a number of ways: by ignoring records, by manually filling in missing values or using global constants, or by using attribute means or most likely values through inference based on Bayesian formulas or decision trees. Missing values were observed when certain stocks or indices were recorded as null on certain dates, and to appropriately reduce the workload of this study, only stocks with up to ten null values were selected for processing, which included skipping entire records containing missing values, or filling missing values with the mean, or using inference (e.g., based on most similar instances). All null values appearing in this experiment are filled using the average of their closest left and right nonnull values, making the stock time series data complete.
- (2) *Data Noise Reduction.* Data contain a certain amount of noise due to the complexity of market dynamics. Noisy data are attributed to random errors or variations in measured variables, as well as errors in data collection tools, data entry errors, etc. The noise present in stock datasets can usually be classified into three main categories: duplicate records, inconsistent stock names, and incomplete files. Some files contain duplicate records. Therefore, all files are thoroughly checked to ensure that only unique records are available on the same date. When

a public company changes its name or a company merges, there may be inconsistencies in stock names. In this case, some of the data can use the old name while the rest of the data can use the new name. The standard is set in solving this inconsistency problem, and all the old names appearing in the records are modified in bulk with the updated records. Noise reduction is done by pywt library wavelet transform.

- (3) *Data Normalization*. Normalization is the “scaled down” transformation of data. If the value of one component of an attribute is too large, the other attributes may lose their moderation effect. For example, when stock price and trading volume are taken together as characteristic values, the difference between the two in terms of data volume is great because the stock trading volume is huge and can reach the level of billions, while the stock price is only a few tens or hundreds, but it cannot be shown that the impact on price is proportionally greater because the value of the trading volume is larger. The convergence of the model is often affected because the data scale is not consistent, the gradient is not “uniform,” the model is not stable when training, and the gradient descent algorithm does not easily converge to the optimal solution. Therefore, in this paper, the data are normalized. When normalization is performed, the size of all values is scaled to a fairly low value. The two most common methods of data normalization are min-max normalization and  $z$ -score normalization.

In order to improve the training efficiency of the samples and the generalizability of the training results, this paper uses min-max normalization to normalize the input feature sequences by which the data inputs are mapped to a predefined range  $[0, 1]$  or  $[-1, 1]$ . The min-max method normalizes the value of attribute  $A$  of the dataset according to the minimum and maximum values of the dataset. It converts the value  $a$  of attribute  $A$  to  $a$  in the range  $[\text{low}, \text{high}]$  by the following calculation:

$$a = \frac{\text{high} - \text{low}}{\max A - \min A} + \text{low}. \quad (1)$$

In particular, when low is set to 0 and high is set to 1, it is easy to see that  $a = 0$  when  $a = \min$  and  $a = 1$  when  $a = \max$ . This means that the minimum value in  $A$  maps to 0 and the maximum value in  $A$  maps to 1. Thus, the entire range of  $A$  values from the minimum to the maximum value maps between 0 and 1.

The data is then subjected to word separation processing. Since Chinese word separation techniques are relatively mature and are not the focus of this thesis, the existing jieba word separation technique, which is suitable for large-scale text separation scenarios, is used here. The TF-IDF method is a common word weighting measure that indicates the importance of a particular word in the whole document or

corpus, and the more frequently a word appears in a document to which it belongs, but hardly appears in all documents, the more it represents the key content. The importance of each word is determined by counting the total number of financial commentary documents, the number of document words, the number of times the word appears in a particular document, and the number of times the word appears in all documents. The formula for calculating the importance of a word in a given stock forum comment document is

$$T_f = \sum_{i=1}^n w_i + \text{ID}_F. \quad (2)$$

TF stands for word frequency and gives the frequency of words in each document in the corpus. It is calculated by the ratio of the number of occurrences of a word in a document to the total number of words in that document. It increases as the number of occurrences of the word in the document increases. Each document has its own TF. The formula is as follows:

$$T_f = \frac{n_{i,j}}{\sum_{i=1}^n w_i}. \quad (3)$$

$\text{ID}_F$  stands for inverse data frequency and is used to calculate the weights of rare words in all documents in the corpus. When  $T_f$  measures word frequency, the weight of words such as “of” or “and” must be reduced since they occur frequently in all documents, i.e., the document inversion frequency component. If a word appears frequently in each document, the less likely it is to be used as a keyword for a given document. The inverse document frequency of a word in a set of documents means how common or rare the word is in the entire set of documents. Thus, if the word is very common and occurs in many documents, the number will be close to 0. Otherwise, it will be close to 1. Designed to retain distinctive words as markers, words that occur rarely in the corpus have a high  $\text{ID}_F$  score. This metric can be calculated by dividing the total number of documents by the number of documents containing the word and then calculating the logarithm. It is given by the following formula:

$$\text{ID}_{F,n} = \log \left( \sum_{i=1}^n w_i \right). \quad (4)$$

Combining these two,  $\text{ID}_{F,n}$  extracts high-frequency words from the text as candidate keywords and the text inverse frequency  $\text{ID}_F$  weights the  $T_f$ - $\text{ID}_F$  scores of the words in the documents in the corpus  $w$ . Multiplying these two numbers will give the  $T_f$ - $\text{ID}_F$  scores of the words in the documents. The higher the score, the more relevant the word is in that particular document, i.e., the one with the higher weight is taken as the keyword:

$$w_{i,j} = T_{f,i,j} + \log \left( \sum_{i=0}^n w_i \right), \quad (5)$$



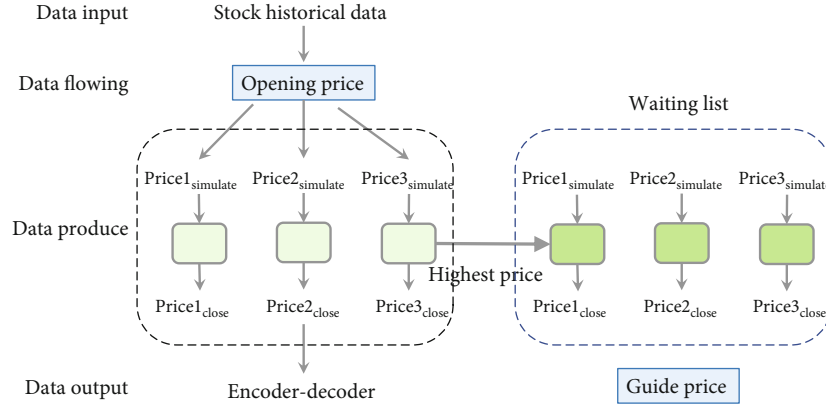


FIGURE 2: Deep recursive process.

where  $TF_{i,j}$  denotes the number of occurrences of  $i$  in  $j$  texts,  $df_i$  denotes the number of corpora containing  $i$  words, and  $N$  denotes the number of all corpora. Regarding the embedding layer, this model maps each word in a set  $K$  of size  $n$  to a corresponding word vector  $w$  by using the Word2Vec method. For the set of vectors  $w$ , the information is extracted using a fully connected neural network:

$$\text{Word}_2\text{Vec} = a\bar{w} + R_n, \quad (6)$$

where  $a$  is the weight size of each word vector,  $a$  belongs to  $R_n$ ,  $b$  is the weight size of the bias vector  $b$ ,  $b$  belongs to  $R_n$ , and  $\text{New}$  is the news information extraction result vector.

In order to correspond the stock names to the news data, each stock name is transformed into the corresponding stock vector  $S$  by embedding operation. By the Word2Vec method, it is transformed into the corresponding word vector  $c_v$ . The generated word vector is fed into CLSTM, and the stock name Name is used as Topic to process the stock news keyword Keys; in particular, the output matrix of CLSTM is used here as output to obtain the corresponding implied layer matrix information  $c_h$ :

$$c_h = \text{clastm}(\text{name}, \text{keys}). \quad (7)$$

**2.1.3. Deep Recursive Process.** The input layer of the model includes three major parts: stock historical data information, news information, and investor sentiment. The main part of the model firstly uses Bi LSTM to process the features in terms of data information separately, secondly uses news data as the contextual information input of CLSTM, and introduces the attention mechanism to give different attention to the news text sequences to ensure that the model captures information from the news that is more relevant to the stock price movement. Secondly, a plain Bayesian-based sentiment classifier is used to analyze the data from the forum, and the sentiment data obtained from it is combined with other parts of the data to serve as training data for the long- and short-term memory time series learning model. Finally, a multilayer fully connected neural network is used to process all the data. Satisfactory results can be obtained

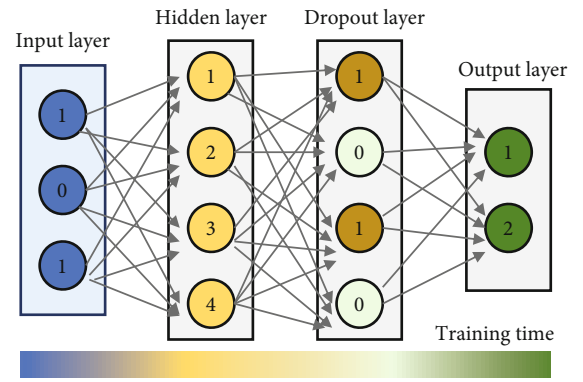


FIGURE 3: Schematic of adding dropout neural network.

by training the neural network, as shown in Figure 2. It also takes into account the effect of some randomness to avoid the occurrence of falling into a local minimum that leads to a global minimum not being reached.

The type of optimizer used optimizes the efficiency with which the algorithm converges to a minimum. The model is optimized with the Adam optimizer, which combines the advantages of both ADAGRAD and RMSPROP optimizers. The reason behind ADAGRAD is that infrequently used parameters must have a large learning rate, while frequently used parameters must have a small learning rate. The stochastic gradient descent update of ADAGRAD becomes

$$\beta_{t+1,i} = \beta_{t,i} - \eta, \quad (8)$$

$$\eta_{t,i} = \nabla J_{i,j} K. \quad (9)$$

The learning rate is derived by calculating the historical gradient of each parameter. Thus,

$$\beta_{t+1} - \beta_t = \frac{\theta}{\sqrt{Q + \chi}} h_t, \quad (10)$$

where  $H$  is the sum-of-squares matrix of the historical gradients. The problem with this optimization is that as

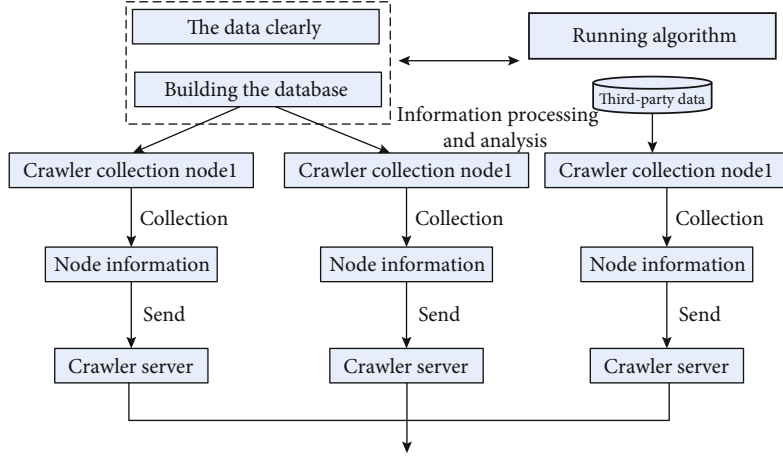


FIGURE 4: Web crawler flow.

the number of iterations increases, the learning rate begins to rapidly decrease and disappear. RMSprop slows down the decline in the learning rate by using a certain number of historical gradients. Updated to

$$\beta_{t+1} - \beta_t = \frac{\theta}{\sqrt{Q\sqrt{[t^2 - r]}}} h_t, \quad (11)$$

$$Q\sqrt{[t^2 - r]} = Q\sqrt{t} + 0.1t^2. \quad (12)$$

After using two hidden layer neural networks in this model, as shown in Figure 3, two hidden layer standard neural networks are used, and the dropout method is adopted to randomly discard some neurons, and the neurons marked with a fork in the right figure are randomly inactivated neurons. In practice, the dropout value is set too low, and the effect can be ignored; if the dropout value is set too high, it may lead to underfitting results. Therefore, the dropout is set to 20% in this model.

## 2.2. Experiments and Result Analysis

**2.2.1. Experimental Data.** Stock forums provide a platform for investors to discuss online. Stock forum sites allow users to request and exchange information therein. In addition, the stock forum site also allows users to view forum posts and post messages therein. When posting a message, users can create a new topic or post in an existing topic. Stock forums consist of user-generated content (UGC), and to get investor sentiment from a forum site, its content should be downloaded first. Use web crawler technology to crawl content and obtain data. A web crawler, an important part of current search engines, is a computer program or automated script that automatically crawls and downloads web information according to certain rules. In a narrow sense, it is usually considered a software program that traverses the information space of the World Wide Web based on web hyperlinks and web document retrieval methods (e.g., depth-first or breadth-first) using the standard HTTP protocol. Web crawlers obtain the content of the web pages corre-

sponding to each URL by determining the queue of URLs to be crawled, parse the web page content, and store the corresponding data. In order to efficiently and accurately obtain comment data, the crawler should start from the entry URL of the forum. Web crawlers are classified into several categories according to the system structure and technical mode, such as general-purpose web, focused web crawlers, incremental web crawlers, and deep web crawlers. In this paper, we use a general-purpose web crawler to crawl the stock forum comment data. It is very expansive, crawling objects involving URLs all the way to the entire Web, collecting data for web search engines and large web service providers. Due to the large scope and number of crawls, high requirements for crawl speed and storage space, and low requirements for the order of crawled pages, usually web crawlers work in parallel. Therefore, the application value is strong. The specific flow chart of the generic web crawler operation is shown in Figure 4.

**2.2.2. Evaluation Based on Accuracy and Yield.** The training performed on the above data is used to predict the stock trend of that day. After trying methods such as fitting stock movements and classifying stock movements, experiments were conducted using the classification method, which is a dichotomous classification (up/down) of stock movements to make predictions. The experiments show that reducing the problem to a classification problem is more accurate than fitting stock movements. To evaluate the impact of financial news on stock prices over time, we set different time intervals (i.e., 1, 2, 5, 7, and 10 days) for the prediction experiments. In the case of a one-day interval, it means that news affects stock prices within 24 hours. Similarly, the impact of news varies at other time intervals. The results are shown in Figure 5, where the accuracy obtained by the prediction model proposed in this paper decreases with time, with the highest accuracy of the experimental results in the first 24 hours and decreases with time. This also illustrates the impact of financial news and the rapid reflection of the stock market.

The experiment uses the above stock selection strategy to backtest the test set by selecting one hundred consecutive

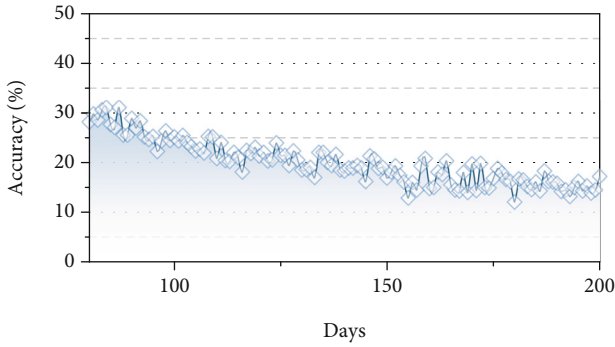


FIGURE 5: Accuracy vs. number of days line graph.

days of data and removing the cases where the data on the day of backtest is less than 100 shares and conducting the test using the return as a measure, the results of which are shown in Figure 6. Daily return represents the relationship between the return available for each investment and the number of days, where the horizontal coordinate indicates the number of days and the vertical coordinate indicates the return, and the points in the figure are the return for each investment with the dashed 0 coordinate, where the number of days with a return greater than 0 is  $G_{days}$ . The results indicate that the actual probability of a daily return greater than 0 is close to 0.7041.

Figure 7 shows the return versus time assuming that the investment strategy given by the model is adopted every day, where the horizontal coordinates indicate the number of days and the vertical coordinates indicate the return compared to day 0, where the coordinates of the highest point are (98, 0.2849), i.e., if the stock selection is done with this model, then at 98 days, a cumulative return equal to 28.49% of the principal can be obtained.

The number of input features of the network model is adjusted, and the same training data and test data are experimented with in a single LSTM model, a news-based LSTM model, and a news- and investor sentiment-based LSTM model, and the results are shown in Figure 8. In the training process of the model, the accuracy of stock trend prediction by LSTM with a different number of layers is compared by continuously modifying the LSTM layers. However, the number of layers is positively correlated with the computational redundancy and consumption, which is not conducive to the efficiency of the overall model. Therefore, with the proper number of network layers, increasing the number of network layers is costly and the improvement of prediction accuracy is not significant. Therefore, a two-layer LSTM network model is used to predict stock movements.

In this paper, stocks and their corresponding news are extracted to obtain information about possible stock trends from the news, such as national policy support that may lead to a rise in the stock. Since each news item is too long, thus, TF-IDF, a common weighting technique used for information retrieval and data mining, is used to process each news item and extract the news-related keywords as input. In order to align with numerical data, historical data from November 5, 2018, to November 8, 2019, were collected,

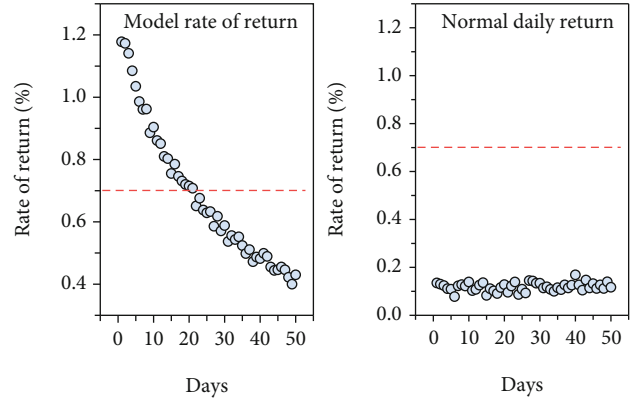


FIGURE 6: Daily yield.

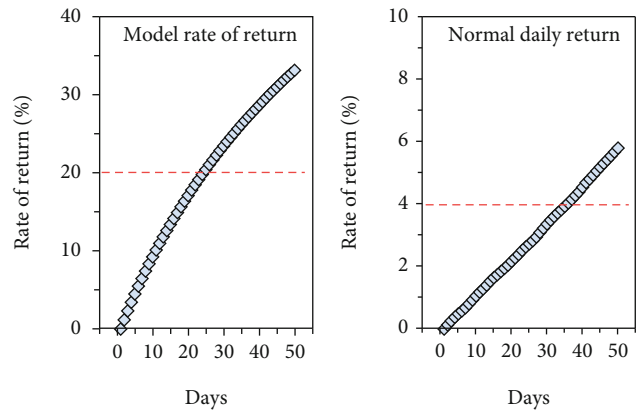


FIGURE 7: Total return.

along with the corresponding stock name information as input, and according to its news release time, 4:00 p.m. was used as the dividing line, with news released before 4:00 p.m. counted as financial news of the day and news released after 4:00 p.m. counted as the next trading day. Also, obtaining stock forum data relies on web crawlers. A plain Bayesian classifier trained using the method described in Section 3 is used to obtain investor sentiment for predicting stock movements. After the above experiments, we verified the conjecture that when the attention mechanism of the hybrid neural network assigns a weight to the  $K$ -line pattern of a sample that is significantly higher than the overall mean of the sample, it is highly likely that the  $K$ -line pattern carries a stronger predictive power than the general  $K$ -line pattern, and when the pattern appears, there is a high probability that the later market will appear to be consistent with the history. It is also true that  $K$ -line patterns are an intuitive analysis tool, and when they show a strong trading signal, human researchers will also focus on the signal to make predictions about the future of the market. Therefore, compared to other deep learning models, the hybrid neural network model proposed in this paper guarantees the predictive effect and performs the predictive inference work in a way that humans can intuitively feel and understand. In addition,

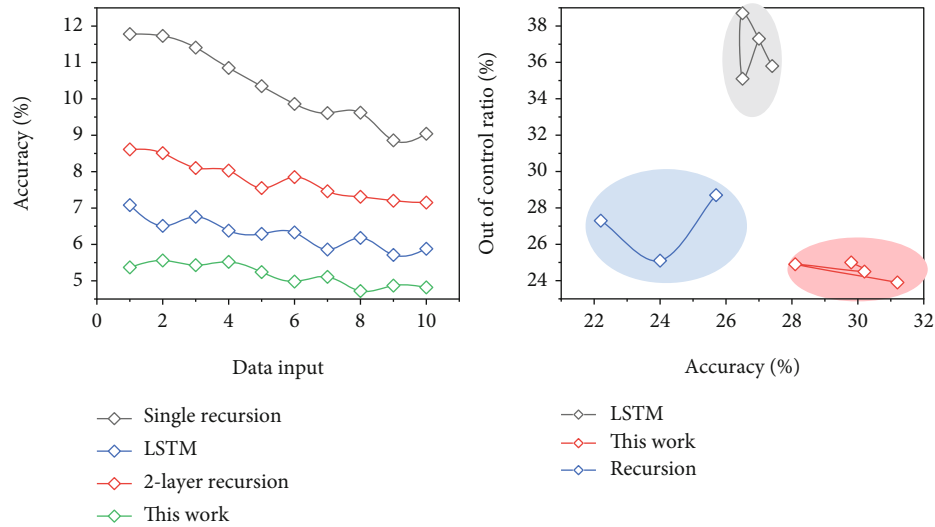


FIGURE 8: Comparison of the accuracy of different prediction models.

using the powerful computing power and storage capacity of computers, we can use the model to mine more  $K$ -line patterns that may have strong predictive power and continuously expand the knowledge base of human researchers.

### 3. Conclusion

This paper focuses on the problem of stock trend prediction based on deep learning. Due to the many factors affecting the stock movement, the number of stocks and the huge trading volume make this research challenging and difficult. In this paper, we consider multiple influencing factors at the same time, combining three aspects: historical stock trading data, news information, and investor sentiment index, and use an improved LSTM to model the prediction. We also construct a stock trend prediction system and apply the improved model to make stock recommendation and stock trend prediction for investors. It has high operability and practical value. In this paper, we classify the sentiment of stock forum remarks and calculate to obtain the investor sentiment index. Based on the sentiment dictionary, a stock market sentiment dictionary is formed by adding stock market specialized vocabulary. A sentiment classification model is constructed on the basis of the plain Bayesian algorithm to classify the sentiment of stock forum remarks. Experiments show that the classification effect is better using this model, and the investor sentiment is obtained more accurately, and the accuracy rate can reach 85%, which lays the foundation for the establishment of the whole stock trend prediction model. In this paper, we design and implement a deep neural network-based stock trend prediction system and upload the trained prediction model to the stock prediction module. Through the requirement analysis of the stock trend prediction system and the design of each functional module, the whole stock trend prediction system is completed after development and testing. It enables investors to make relatively correct investment decisions based on the stock recommendation results and stock prediction results of the

system, thus reducing the possible investment risks to a certain extent and obtaining high and stable investment returns.

### Data Availability

The data used to support the findings of this study are included within the article.

### Conflicts of Interest

All the authors do not have any possible conflicts of interest.

### References

- [1] M. Wen, P. Li, L. Zhang, and Y. Chen, "Stock market trend prediction using high-order information of time series," *IEEE Access*, vol. 7, pp. 28299–28308, 2019.
- [2] W. Chen, M. Jiang, W.-G. Zhang, and Z. Chen, "A novel graph convolutional feature based convolutional neural network for stock trend prediction," *Information Sciences*, vol. 556, pp. 67–94, 2021.
- [3] M. Z. Asghar, F. Rahman, F. M. Kundi, and S. Ahmad, "Development of stock market trend prediction system using multiple regression," *Computational and Mathematical Organization Theory*, vol. 25, no. 3, pp. 271–301, 2019.
- [4] H. Ni, S. Wang, and P. Cheng, "A hybrid approach for stock trend prediction based on tweets embedding and historical prices," *World Wide Web*, vol. 24, no. 3, pp. 849–868, 2021.
- [5] F. Zhou, H.-m. Zhou, Z. Yang, and L. Yang, "EMD2FNN: a strategy combining empirical mode decomposition and factorization machine based neural network for stock market trend prediction," *Expert Systems with Applications*, vol. 115, pp. 136–151, 2019.
- [6] M. Zhao, A. Jha, Q. Liu et al., "Faster mean-shift: GPU-accelerated clustering for cosine embedding-based cell segmentation and tracking," *Medical Image Analysis*, vol. 71, article 102048, 2021.

- [7] A. Moghar and M. Hamiche, "Stock market prediction using LSTM recurrent neural network," *Procedia Computer Science*, vol. 170, pp. 1168–1173, 2020.
- [8] M. Zhao, Q. Liu, A. Jha et al., "VoxelEmbed: 3D instance segmentation and tracking with voxel embedding based deep learning," 2021, <https://arxiv.org/abs/2106.11480/>.
- [9] J. Zhang, S. Cui, Y. Xu, Q. Li, and T. Li, "A novel data-driven stock price trend prediction system," *Expert Systems with Applications*, vol. 97, pp. 60–69, 2018.
- [10] A. Thakkar and K. Chaudhari, "CREST: cross-reference to exchange-based stock trend prediction using long short-term memory," *Procedia Computer Science*, vol. 167, pp. 616–625, 2020.
- [11] W. Sun, P. Zhang, Z. Wang, and D. Li, "Prediction of cardiovascular diseases based on machine learning," *ASP Transactions on Internet of Things*, vol. 1, no. 1, pp. 30–35, 2021.
- [12] M. Nabipour, P. Nayyeri, H. Jabani, S. Shahab, and A. Mosavi, "Predicting stock market trends using machine learning and deep learning algorithms via continuous and binary data; a comparative analysis," *IEEE Access*, vol. 8, pp. 150199–150212, 2020.
- [13] M. Vijh, D. Chandola, V. A. Tikkiwal, and A. Kumar, "Stock closing price prediction using machine learning techniques," *Procedia Computer Science*, vol. 167, pp. 599–606, 2020.
- [14] S. Paul and S. Vishnoi, "Real-time stock trend prediction via sentiment analysis of news article," *Computer Engineering and Intelligent Systems*, vol. 9, no. 7, pp. 21–28, 2018.
- [15] T. Kim and H. Y. Kim, "Forecasting stock prices with a feature fusion LSTM-CNN model using different representations of the same data," *PLoS One*, vol. 14, no. 2, article e0212320, 2019.
- [16] W. Long, L. Song, and Y. Tian, "A new graphic kernel method of stock price trend prediction based on financial news semantic and structural similarity," *Expert Systems with Applications*, vol. 118, pp. 411–424, 2019.
- [17] M. A. I. Sunny, M. M. S. Maswood, and A. G. Alharbi, "Deep learning-based stock price prediction using LSTM and bi-directional LSTM model," in *2020 2nd Novel Intelligent and Leading Emerging Sciences Conference (NILES)*, pp. 87–92, Cairo, Egypt, 2020, October.
- [18] H. S. Ibrahim, A. Z. Khan, S. Attia, and Y. Serag, "Classification of heritage residential building stock and defining sustainable retrofitting scenarios in Khedivial Cairo," *Sustainability*, vol. 13, no. 2, p. 880, 2021.
- [19] T. Colliri and L. Zhao, "Stock market trend detection and automatic decision-making through a network-based classification model," *Natural computing*, pp. 1–14, 2021.
- [20] Y. Gu, A. Chen, X. Zhang, C. Fan, K. Li, and J. Shen, "Deep learning based cell classification in imaging flow cytometer," *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 2, pp. 18–27, 2021.
- [21] Y. Yan and D. Yang, "A stock trend forecast algorithm based on deep neural networks," *Scientific Programming*, vol. 2021, Article ID 7510641, 7 pages, 2021.
- [22] Z. Wang, P. Zhang, W. Sun, and D. Li, "Application of data dimension reduction method in high-dimensional data based on single-cell 3D genomic contact data," *ASP Transactions on Computers*, vol. 1, no. 2, pp. 1–6, 2021.
- [23] J. P. Verma, S. Tanwar, S. Garg, I. Gandhi, and N. H. Bachani, "Evaluation of pattern based customized approach for stock market trend prediction with big data and machine learning techniques," *International Journal of Business*, vol. 6, no. 3, pp. 1–15, 2019.
- [24] Y. Chen, W. Lin, and J. Z. Wang, "A dual-attention-based stock price trend prediction model with dual features," *IEEE Access*, vol. 7, pp. 148047–148058, 2019.
- [25] M. Y. Chen, C.-H. Liao, and R.-P. Hsieh, "Modeling public mood and emotion: stock market trend prediction with anticipatory computing approach," *Computers in Human Behavior*, vol. 101, pp. 402–408, 2019.
- [26] M. Ananthi and K. Vijayakumar, "Stock market analysis using candlestick regression and market trend prediction (CKRM)," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 5, pp. 4819–4826, 2021.
- [27] S.-F. Huang, M. Guo, and M.-R. Chen, "Stock market trend prediction using a functional time series approach," *Quantitative Finance*, vol. 20, no. 1, pp. 69–79, 2020.
- [28] L. Huang, G. Xie, W. Zhao, Y. Gu, and Y. Huang, "Regional logistics demand forecasting: a BP neural network approach," *Complex & Intelligent Systems*, pp. 1–16, 2021.
- [29] Y. Fang, "Feature selection, deep neural network and trend prediction," *Journal of Shanghai Jiaotong University (Science)*, vol. 23, no. 2, pp. 297–307, 2018.
- [30] C.-s. Lim, "Fair performance evaluation method for stock trend prediction models," *The Journal of the Korea Contents Association*, vol. 20, no. 10, pp. 702–714, 2020.
- [31] L. Lei, "Wavelet neural network prediction method of stock price trend based on rough set attribute reduction," *Applied Soft Computing*, vol. 62, pp. 923–932, 2018.
- [32] C. Patel, "Impact of the financial result on stock trend prediction using SVM and M-score," *International Journal for Advance Research and Development*, vol. 3, no. 5, pp. 6–11, 2018.
- [33] A. Thakkar and K. Chaudhari, "A comprehensive survey on portfolio optimization, stock price and trend prediction using particle swarm optimization," *Archives of Computational Methods in Engineering*, vol. 28, no. 4, pp. 2133–2164, 2021.
- [34] C.-s. Lim, "A study on stock trend determination in stock trend prediction," *Journal of the Korean Society of Computer and Information*, vol. 25, no. 12, pp. 35–44, 2020.
- [35] V. Gupta and M. Ahmad, "Stock price trend prediction with long short-term memory neural networks," *International Journal of Computational Intelligence Studies*, vol. 8, no. 4, pp. 289–298, 2019.

## Research Article

# Research on the Role of Big Data Technology in the Reform of English Teaching in Universities

Xiaoge Jia 

*Department of Foreign Languages, Xijing University, Xi'an 710123, China*

Correspondence should be addressed to Xiaoge Jia; [lois888777@163.com](mailto:lois888777@163.com)

Received 11 July 2021; Revised 1 August 2021; Accepted 16 August 2021; Published 13 September 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Xiaoge Jia. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper provides an in-depth understanding and analysis of the reform of English teaching in colleges and universities by analyzing the role of big data technology for the reform through in-depth research and analysis. Based on the background of the era of education informatization, this study explores the transformative value of the integration of information technology and teaching activities and elaborates the relevant significance at the theoretical and practical levels. Based on this, the research method of this study is established. The realm of integration of information technology and teaching activities and its transformative value is taken as the focus of the study. The integration of information technology and teaching activities means that it is not only a simple superposition between information technology and teaching but also a process to explore the role and influence of information technology into teaching activities by deeply exploring the inner connection between the two, to complete the integration of information technology and teaching activities, and finally to realize the comprehensive development of student's personality.

## 1. Introduction

Along with the high speed of information dissemination and rapid economic development, computer Internet technology has been very commonly used in many fields, computational science entered the field of education, and the use of computer science in education was a new change achieved for traditional education [1]. The application of computers in pedagogy can also be referred to as Computer-Based Education (CBE). Computer-assisted education refers to the application of computers, multimedia, and other technologies in the field of education and teaching to assist in education and teaching. It is a combination of traditional education methods and rational use of educational resources to improve education standards. Traditional teaching has become outdated; the integration of technology and classroom has brought significant changes to education, and today, we are entering the era of big data, and data mining has been widely used in various industries to extract valuable knowledge and provide decision support [2]. This technology has also been gradually integrated into the field of education, opening a new horizon for the development of smart class-

rooms [3]. Technological advances and multimedia applications have brought opportunities for the transformation of traditional classroom teaching, and the rapid development of education informatization has provided opportunities for in-depth research on smart education and smart classroom, and smart education has become a hot topic in the development of education informatization [4]. With the progress of society and the development of science and technology, the application of information technology in the field of education has become extensive, and the development of smart education is the new form and requirement of the development of information technology, education informatization, and education reform in modern society. The arrival of the era of big data has driven the development of all occupations, and the field of education is also being deeply affected, and the penetration of advanced big data technology into the field of education is the trend of the future development of education informatization, which is bound to have a great impact on the profound changes in the field of education [5].

Innovation of smart classroom teaching mode is one of the important strategies related to the rapid progress of China's education, cultivating innovative talents, and

changing education mode, and it is also an inevitable trend to enhance teaching effect [6, 7], improve teaching quality, perfect teaching system, and promote education development, and it is about to become a trend of the times to put education development in the big data environment. This study focuses on the deep integration of new technology and teaching in the new era, explores the teaching model of the smart classroom under the big data environment, designs relevant smart classroom teaching cases, and implements them, which is of great significance to the innovation of teaching mode in the education field and has certain theoretical and practical significance to the development of smart education. At present, personalized learning has been effectively realized with the development of artificial intelligence, and it is obvious that artificial intelligence plays an important role in the fields of learning tutoring, learning assessment, and teaching optimization, which promotes the all-round improvement of teaching efficiency and better learning experience for students. Artificial intelligence [8–10] has gradually become an important factor in the development of education information technology, helping the innovative development of education and teaching. The development of information technology has not only brought a huge impact on the field of science and technology but also has an important role in the development of times that cannot be ignored. At present, the education field is actively introducing artificial intelligence technology, and the elements of teaching activities are constantly changing, promoting students' learning activities toward personalization and lifelong learning and pushing the whole education level from low-level crude education to high-level precise education.

This study analyzes the new educational technology of artificial intelligence and the impact of the pedagogical changes that will occur, as well as the basic characteristics of artificial intelligence its significance and impact on the pedagogical aspects, and explores the new concepts of the change of the elements of teaching activities and generation in the era of artificial intelligence. It is also helpful to explore the changes in teaching in the era of AI to expand teaching ideas. Therefore, it is of practical guidance in the future application of AI to optimize teaching methods and teaching techniques in the teaching process and students' learning methods. The learning platform of various online teaching modes provided by artificial intelligence technology is profoundly affecting the development of teaching as a process. The teaching in the era of artificial intelligence will overturn the traditional teaching model, the development of artificial intelligence enlightens the new ideas of teaching change development, and the teaching change profoundly affects the new direction of the future development of the young generation. In traditional education, the main channel for students to acquire knowledge is through teachers' lectures. However, in the era of artificial intelligence, human beings will need to reexamine the ways of acquiring knowledge to adapt to social changes, and the development of technology profoundly affects the change of teaching. This study analyzes the changes in teaching and learning in the age of artificial intelligence to provide a better understanding of teaching and learning in the age of intelligence. The smooth

implementation of intelligent teaching requires the application of artificial intelligence in teaching as a condition. Exploring the changes in the elements of teaching activities and new concepts in the age of artificial intelligence can make personalized teaching and adaptive learning possible in the age of artificial intelligence, and this teaching method can optimize the traditional teaching methods, enhance teaching efficiency, improve teaching outcomes, enhance teaching participation, and promote teaching activities in the direction of intelligence, precise, and personalized direction.

## 2. Current Status of Research

The impact of AI technology on educational activities has become increasingly evident, and research on AI in education has become a focus of foreign academics and has yielded a variety of research results. Hasan et al. suggested that AI in education is a completely new research area, and this type of research is conducted on the premise that educational activities and artificial intelligence are closely linked [11]. Aho and Duffield suggested that educational AI is mainly used to improve students' participation in teaching activities and teachers' teaching efficiency with the help of the Internet teaching platform [12]. In the future, the focus of educational AI will be on the development of intelligent robots, which will try to improve students' learning efficiency by adding intelligent sensing devices. In addition, some scholars have conducted in-depth studies on the problems that arise after the application of AI in education to study the problems that may arise after the application of AI to education and the future development trend [13]. With the continuous development of AI technology, many researchers are now studying the design of machine learning algorithms intending to improve the efficiency of their interaction with humans [14]. Modern information technology has brought many changes to the development of education, and at the same time, the impact of changes in information technology has gradually risen to the national level [15]. According to the educational development planning documents released by the education department in China, information technology will play an important role in influencing the development of education, and information technology needs to lead education to modernization and drives changes in education [16].

Knox firstly studied in depth the factors related to the change of teaching methods and made a detailed analysis of the direction of the change of teaching methods, pointing out that the change in teaching methods is not to negate the past teaching traditions but to optimize and improve the teaching methods based on absorbing the advantages of traditional teaching activities, and the scholars also studied the direction of the reform of teaching models, which believed that the reform of teaching models would mainly focus on four directions, such as optimizing teaching resources and innovating teaching organization [17]. Tian et al. designed two types of collaborative learning models, formal and sub-formal, and through the study and analysis of these two models, they proposed that the intelligent terminal devices

and platforms used in the smart classroom can help teachers to grasp students' learning situation and provide rich learning resources to students [18]. It is also possible to provide students with a wealth of learning resources, enabling resource sharing and interoperability between teachers and students [19]. Teacher-student interaction directly determines student learning outcomes, and interactive learning among students promotes achievement and social-emotional development. Council member, after extensive research and analysis, indicated that a high frequency of classroom interactions helps students develop self-confidence in their learning and that students are motivated to learn and actively participate in classroom learning activities [20]. Xiao and Yi showed that teachers who frequently ask challenging questions and provide students with cues other than words are effective in driving students' cognitive development [21].

Firstly, through the analysis of the existing literature, the value orientation of the integration of information technology and teaching activities is clarified, the feasibility of the research problem is determined, and the exact connotation and way of the integration of information technology and teaching activities are determined based on this; secondly, the realm of the integration of information technology and teaching activities is explored from the theoretical level, and the profound influence of information technology on teaching activities is elaborated to provide theoretical support for the practical level; finally, the changes brought about by information technology to teaching activities are identified, and the main characteristics of each change element are sought, based on the classroom analysis. Finally, based on theory and practice, the transformative value of the integration of information technology and teaching activities is analyzed at the practical level, to provide lessons for future classroom teaching.

### 3. Analysis of the Role of Big Data Technology in the Reform of English Teaching in Colleges and Universities

*3.1. Improved Big Data Technology Design.* The big data environment is a branch environment in the information technology environment, in which the smart teaching model is carried out, which is the culminating stage in the development of information technology education while creating fever in the field of education [22]. Technology can realize the communication between human language and machine language; build data models to comprehensively record and collect behavioral data; use data collection, data intelligence analysis, real-time analysis of learning data, real-time pushing of resource data, dynamic feedback results, and other mutual combinations to promote the construction of the big data environment; and complete the construction and implementation of the wisdom classroom model in the new era of the big data environment so that students experience the teaching process interactive [23]. In addition, I will promote the rapid development of education in the direction of informatization and wisdom by making the teaching pro-

cess interactive, intelligent, data-based, and dynamic. The key to completing intelligent learning lies in how to use the terminal tools and big data analysis techniques in the big data environment to grasp the real-time learning dynamics of students, analyze student learning data, create educational big data, provide feedback on teaching effects, make full use of the big data environment to create intelligent education, and complete the intelligent classroom in the big data environment.

Advances in science and technology and multimedia applications have brought opportunities for the transformation of traditional classroom teaching. The rapid development of educational informatization provides an opportunity for in-depth research on smart education and smart classrooms. Smart education has become a hot topic in the development of educational informatization.

The advent of the big data era has caused a phenomenon of data in various forms gradually spreading into life and society, and of course, the education field is also constantly receiving the changes brought by big data. The generation of educational data has become an important asset and resource for the development of education informatization. By bringing together multidimensional, multilevel, and multifaceted teaching data and then comprehensively organizing and analyzing them through data analysis methods, I can analyze teaching behavior in real-time and provide optimal decisions for education and generate smart education. There are many elements involved in wisdom education, and these elements are closely related to each other wisdom teaching, and the implementation of teaching is largely carried out by various advanced technologies, relying on the wisdom teaching environment, and the collision between big data and classroom rubs out the spark of wisdom, which in turn generates wisdom classroom. The input of big data technology is the premise of smart education, and the smart classroom is the environment that provides the possibility and opportunity for smart education with three-dimensional interaction, multidirectional communication, comprehensive attention to students, timely feedback on learning, and reconstructing the teaching model.

$$y'_B = y_A * -v'_{AC} * (t_B + t_A), \quad (1)$$

$$N = \frac{O + \bar{O}}{K}. \quad (2)$$

The processed data can better reflect the number of user clicks on online courses, making the mining of association rules more accurate. One of the characteristics of artificial intelligence is human-machine collaboration. Specifically, human-machine cooperation means that humans and machines work together to play their roles in their respective areas of expertise. The media theory favors the theoretical aspect of the integration of information technology and teaching activities and only considers information technology as the basis of teaching practice [24]. Although it appears in teaching activities as a physical tool or means, it still fails to view the integration of information technology and teaching activities but only from the perspective



that the functions and characteristics of information technology have an impact on teaching activities. Mediation theory is biased toward the practical aspect of the integration of IT and teaching activities, and to a certain extent, it has realized the shift from “theoretical activities focusing on the characteristics of IT to practical operations focusing on IT,” so that IT can participate in solving difficult problems in teaching activities. However, the pursuit of practical operation will lead to procedural and mechanical teaching activities, which will lead to excessive technical rationality and loss of objective judgment on teaching activities. Innovative smart classroom teaching mode is one of the important strategies related to the rapid progress of my country’s education, the cultivation of innovative talents, and the change of education mode.

The relational theory is a paradigm of integration combining theory and practice, emphasizing the symbiotic organic integration of information technology and teaching activities and attacking the deep integration of information technology and teaching activities from a holistic, comprehensive, and dialectical standpoint, which fundamentally shakes the methodological problem of integration of information technology and teaching activities, as shown in Figure 1.

Data mining discovers knowledge in the form of constraints, patterns, laws, concepts, rules, visualizations, etc. The uncovered knowledge can be used by decision-makers or experts to improve the existing knowledge system, and in the field of education, it can be used by learners, administrators, teachers, and other actors to intervene in the process of education and teaching for inappropriate teaching or content that is not suitable for learners’ learning. The main tasks of data mining are twofold: classification-prediction [25, 26] tasks and descriptive tasks [27]. The classification predictive task learns a model from known classified data and uses that model to interpret new unknown classified data to obtain a classification of those data. Depending on the class labels, they are called classification tasks and prediction tasks, respectively. There are many application areas of data mining, and the process of data preparation requires clarifying the requirements, understanding the data background of the relevant domain and the data sources, etc. Individualized teaching and adaptive learning in the era of artificial intelligence can be made possible. This teaching method can optimize traditional teaching methods and improve teaching efficiency.

The data obtained through the data preparation process is the raw data, and in the data selection step, the raw data needs to be filtered to obtain relevant data and samples, i.e., the target data. In this process, the raw data usually needs to be processed using operations in the database. Further processing of the data, there is no shortage of redundant, irrelevant, and interfering data in the target data, which needs to be preprocessed according to the principles of data consistency and completeness. For missing items, the missing data can be restored by statistical methods by time series and known changes as much as possible. Data transformation refers to the reprocessing of data to adapt to database access and reduce the amount of data.

$$\begin{aligned} \text{TSED}(B) &= \sqrt{(x'_B + x'_B)^2 - (y'_B + y'_B)^2}, \\ T_i &= \{(x_1, y_1, t_1), (x_2, y_2, t_3), \dots, (x_m, y_m, t_m)\}. \end{aligned} \quad (3)$$

The key to intelligent learning lies in how to use terminal tools and big data analysis techniques in a big data environment to master the real-time learning dynamics of students, analyze student learning data, create educational big data, and provide evidence-based feedback on teaching effects. In determining data mining objectives, according to the user’s requirement statement, the purpose of data mining is clarified. There are many types of knowledge in the data mining process, and the methods of knowledge discovery used according to different requirements are the same, so determining the mining objectives in this step is to lay the foundation for the next step. Selecting a suitable algorithm according to the mining target determined in the previous step is a key step in the whole data mining. Usually, the algorithms of data mining are classification, association rules, summarization, clustering, etc. The knowledge extracted in the previous step is interpreted, redundant and irrelevant knowledge is removed, and then, the above process is repeated until the user’s needs are met. The final step is the presentation of the knowledge, delivering the results in a way that the user can understand.

$$d_c(T_i, T_j) = \frac{l_{c1}^2 - l_{c2}^2}{l_{c1}^2 + l_{c2}^2}. \quad (4)$$

Association rules refer to two things that are implicitly connected under a rule, and this connection needs to be discovered by data mining algorithms because of its implicit nature. Specifically, an association rule is an implicit rule such as  $X \rightarrow Y$ .  $X$  is called the lead (left-hand-side (LHS)), and  $Y$  is called the successor (right-hand-side (RHS)) of the association rule, and the lead and successor can predict the value of other attributes based on the attribute of one. The association rule algorithm introduces the concepts of support and confidence, the former refers to the probability of containing several items in the dataset at the same time, and the latter refers to the probability of occurrence of  $B$  in the dataset under the condition of occurrence of  $A$ , conditional probability.

$$P(A|B) = \frac{P(AB)}{P(B)}. \quad (5)$$

The interrelationship between support and confidence can be transformed by probability calculations as follows.

$$\text{Conf}(Y|X) = P(Y|X). \quad (6)$$

In addition to the two concepts of support and confidence in association rules, frequent itemset is also an extremely important concept; itemset is a set of items contained in a transaction and its subsets, and frequent itemset

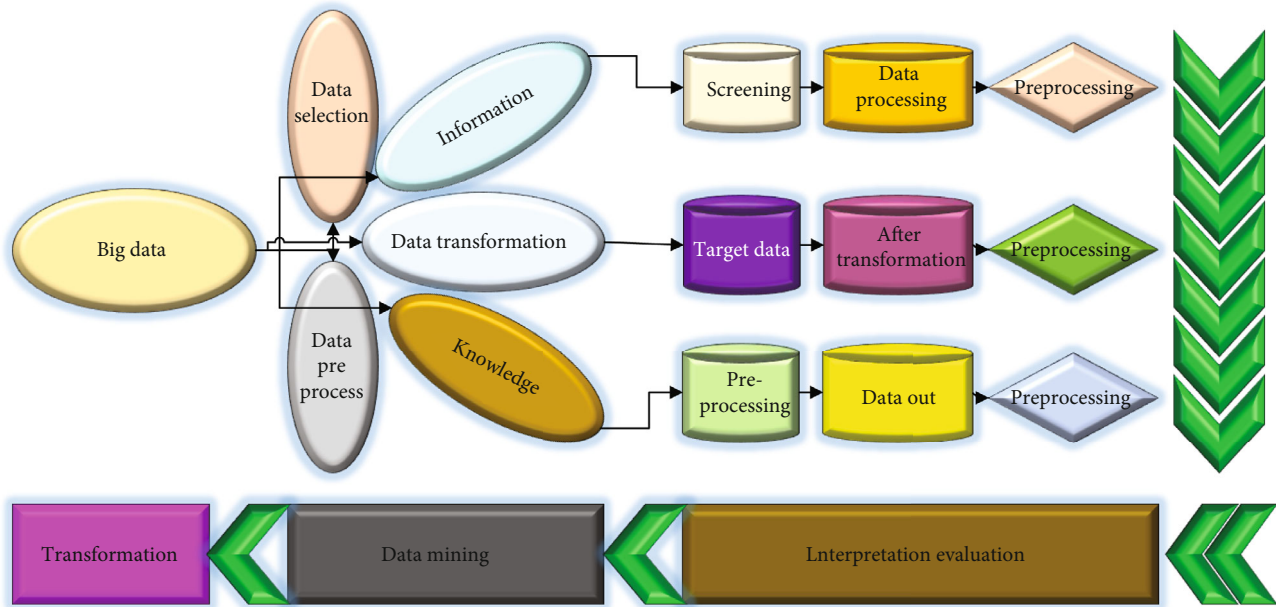


FIGURE 1: Data mining process.

is the set of these items that appear frequently in multiple transactions; the meaning of mining frequent itemsets is to find out the implicit association information between items in a transaction. The significance of mining frequent itemset is to find out the implicit association information between items in a transaction, which is the basis of the association rule data mining algorithm, and the following analysis of common association rule mining algorithms, as shown in Figure 2.

The investment of big data technology is the prerequisite for the development of smart education, and the smart classroom is the environment. It provides the possibility and opportunity for smart education to provide three-dimensional interaction, multidirectional communication, comprehensive attention to students, timely feedback on academic conditions, and reconstruction of teaching models. The processing of data in a complete system is extremely complex and cannot be clearly described by a data flow diagram, and a flow diagram cannot describe multiple data processing at all. The meaning of user data flow diagrams expressing data processing is lost, and there is no way to communicate with users through data flow diagrams. For the data flow diagram to effectively describe the data flow of each functional module, the data flow diagram needs to be divided layer by layer. The data flow diagram is divided into functional modules in a top-down, outside-in, layer-by-layer manner. This can be used to communicate with users or as a basis for developers to design and develop the system.

*3.2. Role Design in the Reform of English Teaching in Higher Education.* After the functional requirements analysis of a system is completed, the basic functions to be performed by the system are defined. However, if you want to ensure that a system performs a specific function in a stable and

orderly manner, you need the system to meet certain non-functional requirements as well. This is because a system can be influenced by the external environment when it is running, and the frequent interactions within the system can also bring unexpected results [28]. The analysis of non-functional requirements of the system is to ensure that the system can run stably and perform the functions provided by the system with quality. First, the level of development of information technology determines the characteristics of students in teaching activities. In the teaching activities, the students' development is partly supported by information technology, which gradually increases from the initial functions of storage and transmission to the functions of processing and even substitution, replacing part of the students' work and to some extent leading to their liberation. There are many application fields of data mining. In the process of data preparation, it is necessary to clarify the requirements and understand the data background and data sources of related fields. Information technology, mainly radio and television, has broken the boundaries of teaching activities, making it possible for all people to learn, and students appear as the public at the social level, while the boundaries of teaching activities in the digital era are more blurred, the requirements for students' information-theoretical level and ability to process information are higher, and the people to be trained in teaching activities are closer to social people. From the perspective of human sociality, the development of students in teaching activities cannot be separated from information technology, whose most basic interaction is mutual dialogue based on language, and it is in the different dialogues that students participate in learning activities and then construct their knowledge structure, to become social people who master knowledge, ability, and literacy. Based on this understanding, after the integration of information technology and teaching activities, the characteristics that

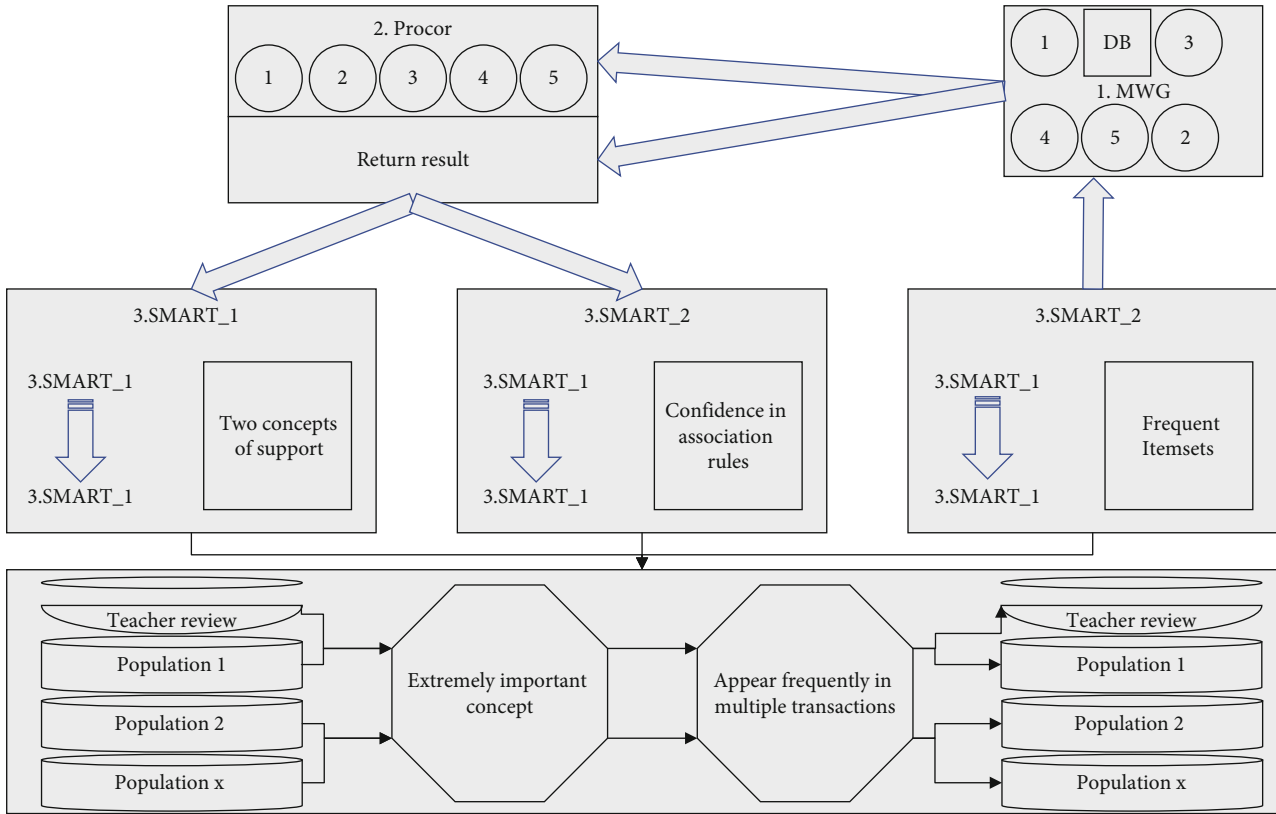


FIGURE 2: Top-level data flow diagram of the system.

students exhibit in teaching activities and the characteristics of the person to be cultivated are closely related to the level and degree of development of information technology, as shown in Figure 3.

The process of integration of information technology and teaching activities is the process of keeping pace with the times, and it is easy to see that with the rapid development of information technology, it is the trend to promote the change of teaching activities with the help of new technology development. The integration of information technology and teaching activities spans two subject areas, namely, information technology and education [29]. The use of information technology leads the development direction of informatization of teaching activities, while the results of informatization in the teaching field also promote the practice and application of information technology. Based on this, the innovative leading characteristics of the integration of information technology and teaching activities are mainly manifested in the following two points. First, innovation leads to the development of information literacy of teachers and students. Since information technology has been introduced into teaching activities, teachers' and students' information literacy has been emphasized. Teachers need to rely on information literacy to interpret information technology knowledge and use information technology equipment, to optimize the teaching process and improve the teaching effect; students need to rely on information literacy to screen the required information, learn to obtain information, and effectively use the obtained information,

to diagnose their learning process and improve the learning effect. Secondly, innovation leads to the development of educational theory and practice. Since the integration of information technology and teaching activities involves a special intersection and integration of disciplines, on the premise that the theory of technical support for teaching and learning is ahead of the updated iterations of technological development, the combing of this lineage not only has a guiding role for future educational practice but also has a model role for the application and promotion of the integration of information technology and teaching activities in the future. This information is the association rule information and is the basis of the association rule data mining algorithm.

However, with the deepening of the integration of information technology and teaching activities, teaching activities become an open whole, and teaching activities no longer contain only unchanging knowledge, but more knowledge is generated by dialogue, cooperation, and interaction between teachers and students through the integration of information technology and teaching activities, which is the product of the interaction between the subject of knowledge and the object of knowledge, and will change with the degree of integration of information technology and teaching activities. Thinking from the perspective of human sociality, the development of students in teaching activities cannot be separated from information technology. The most basic communication is language-based dialogue. It is a product of the interaction between the subject of knowledge and the object of knowledge and will change with the degree

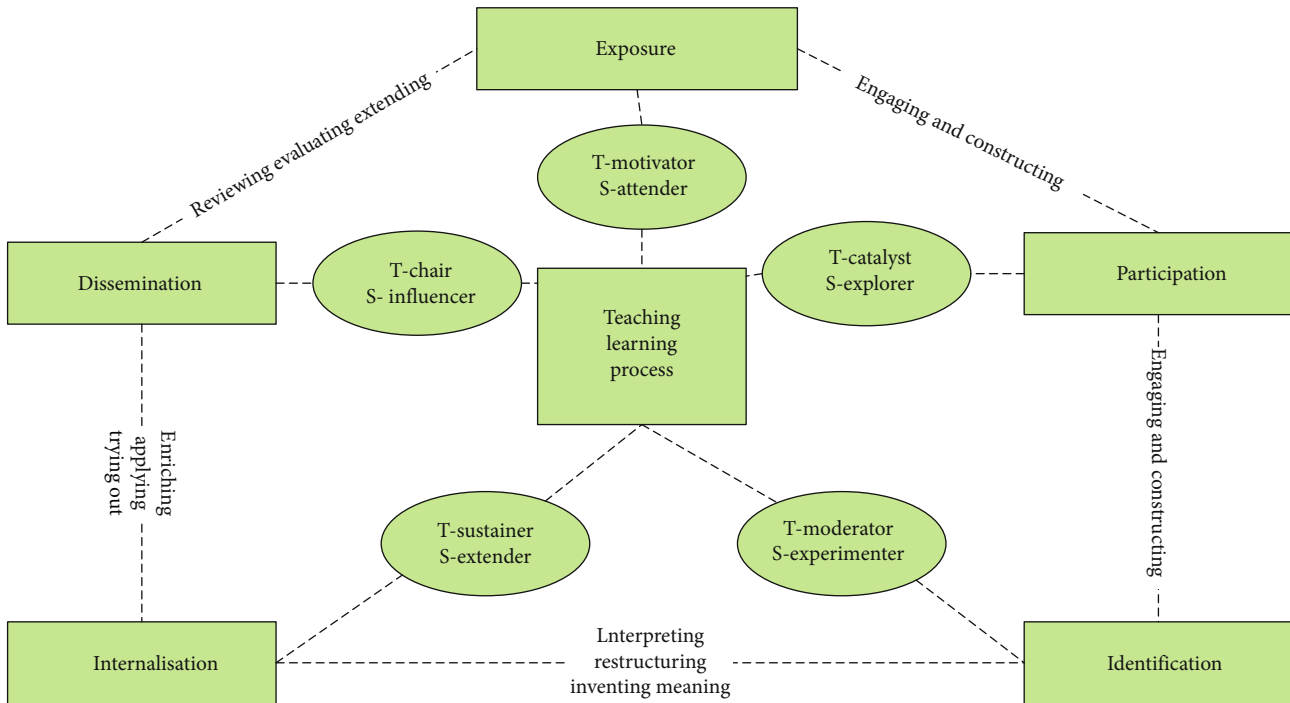


FIGURE 3: Framework of English teaching reform in higher education.

of integration of information technology and teaching activities and is closely related to the inner experience of the subject of knowledge, such as with the deepening of the integration of the two; the knowledge of the subject of knowledge will be continuously improved and accumulated, or the knowledge will be revised. In this process of generating knowledge, knowledge is dynamic and open, constantly changing, and it is not an absolute objective truth, nor is it a closed and stable system.

The postmodern view of knowledge highlights the fluidity of knowledge, advocates the self-construction and generation of knowledge in the process of flow, and emphasizes the openness of knowledge, such as Dewey believes that knowledge is never fixed and eternal; it is the starting point and result of the process of inquiry and as a tool in the process of inquiry can test and confirm the process of practice. In the process of such repeated communication, the knowledge structure of teachers and students will change, and they will integrate their understanding and thinking based on their own experience, to construct the corresponding knowledge. In short, just as the poststructuralist Derrida believed that the real world that people face is a huge system with infinite openness and possibilities, and it does not have the ultimate origin and can be considered as certainty; knowledge is also created through the integration of information technology and teaching activities, which makes the interaction between the subject of knowledge and information technology in the educational field and thus promotes the dynamic process of knowledge construction of the subject of knowledge, as shown in Figure 4.

Through the timely feedback technology in the smart classroom, dynamic evaluation of students' learning process can be realized. In the teaching process, teachers send test

questions online and set the answer time, and students can join the classroom to get the test questions and be asked to turn in the paper within the specified time, which is conducive to the cultivation of students' time concept [30]. With the use of an intelligent learning system, not only can the learning process be transparent and the learning results be visualized, but the teacher can also correct and test the quality of students' learning. After the students' answers are completed, the teacher can analyze the results of the answers to understand the students' learning effects and provide timely feedback to the students, motivate them to learn and make up for their shortcomings with the help of the teacher, and adjust the teaching program based on the students' learning ability. Under the premise that the theory of technical support for teaching and learning is ahead of the update iteration of technological development, the combing of this context not only has a guiding role for future educational practice, but its educational practice exploration will also be used for the application and promotion of the integration of future information technology and teaching activities. It has an exemplary role.

Students are the subjects of learning, and they should also be guided to actively participate in the assessment process. Introducing students into the evaluation system also means that students can self-evaluate, reflect on their successes and shortcomings in learning, discover effective learning methods, and improve their learning ability, which also plays an important role in improving students' personalities. At the same time, students can evaluate each other, supervise each other in mutual evaluation, make progress together, and develop good habits of listening to others' opinions with an open mind, treating others sincerely, and good teamwork spirit. Therefore, we should establish an open and relaxed

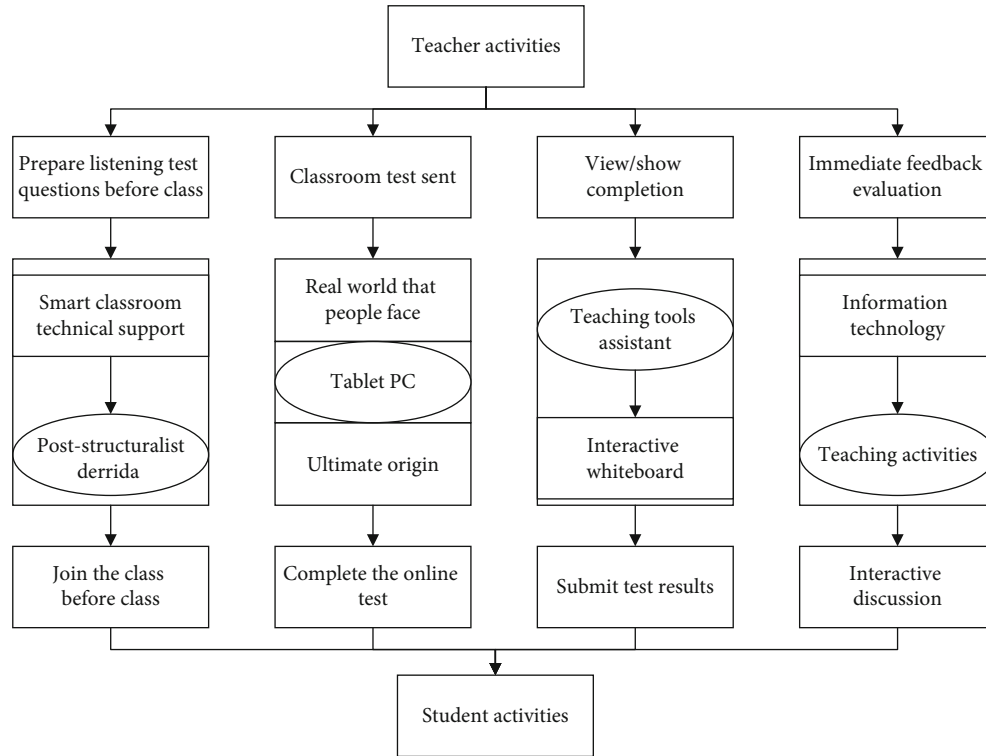


FIGURE 4: Timely feedback process in the classroom.

evaluation atmosphere, focus on teacher evaluation while also encouraging students to participate, realize the diversity of evaluation subjects, and help students to reflect in self-evaluation, teacher evaluation, and mutual evaluation, thus promoting the realization of self-learning and development.

## 4. Analysis of Results

### 4.1. Performance Analysis of Improved Big Data Technologies.

The purpose is to compare the performance of two association rule mining algorithms in a stand-alone environment and analyze the characteristics of the two algorithms to validate the analysis of the two association rule mining algorithms in this chapter, but when the dataset reaches 20,000 items, the execution efficiency of the two algorithms in a stand-alone environment decreases significantly due to the small memory and even fails to run. When the improved association rule algorithm in a parallel environment is used to test the performance of the algorithm on the dataset with less than 20,000 items, it is found that the running time of the algorithm when the data volume is too small is not much different from that of the dataset with 20,000 items, and both are larger than the running time in the stand-alone environment. These are closely related to the inner experience of the cognitive subject. For example, as the degree of integration of the two deepens, the knowledge of the cognitive subject will continue to be improved and accumulated, or the knowledge will be revised. The reason is that since the parallel environment is suitable for large data processing, the algorithm runtime when the data volume is too small is no longer dominated by multinode chunking, and the runtime

is mainly determined by data transfer and copying. The following is the performance analysis of the improved parallel association rule mining algorithm with a parallel environment and a dataset of 20,000 items, as shown in Figure 5.

The number of the initial population is set to 15; the expected number of questions of each type in the test paper is 20, 5, 10, 7, and 5. The maximum number of iterations is 1000, the expected value of fitness is 0.98, the coverage of knowledge points is 0.3, the difficulty coefficient is 0.4, the number of selectors is 10, the number of individuals generated by the cross operator is 20, the total score of each set of questions is 100, the number of questions in the question bank is from 1 to 5000, and the questions in the initial population are selected from 1 to 5000 to combine the questions that satisfy the conditions.

The results of the performance analysis are shown in Figure 6 for 10 experiments conducted under the above initialization conditions with knowledge point coverage of 0.3 and 0.7.

The performance comparison in Figure 6 shows that the average running time of the algorithm is 592.28 ms, and the average number of iterations is 18.4 when the knowledge point coverage is 0.3, and the average running time of the algorithm is 15953.08 ms, and the average number of iterations is 341 when the knowledge point coverage is 0.7; i.e., the performance of the algorithm is improved by 96.29% when the knowledge point coverage is reduced by 40%. The algorithm process shows that when the threshold value of knowledge point coverage is too high, the number of iterations increases significantly, and too low, the quality of the completed paper is affected because the knowledge point

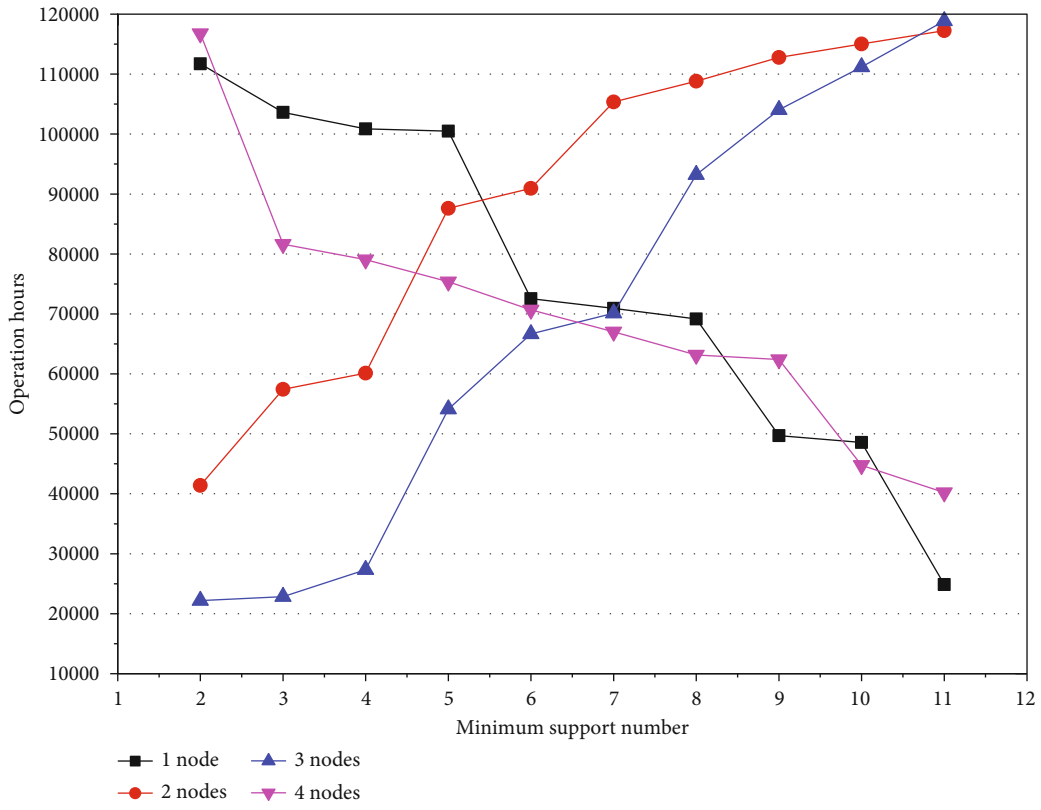


FIGURE 5: Comparison of algorithm efficiency in parallel environments.

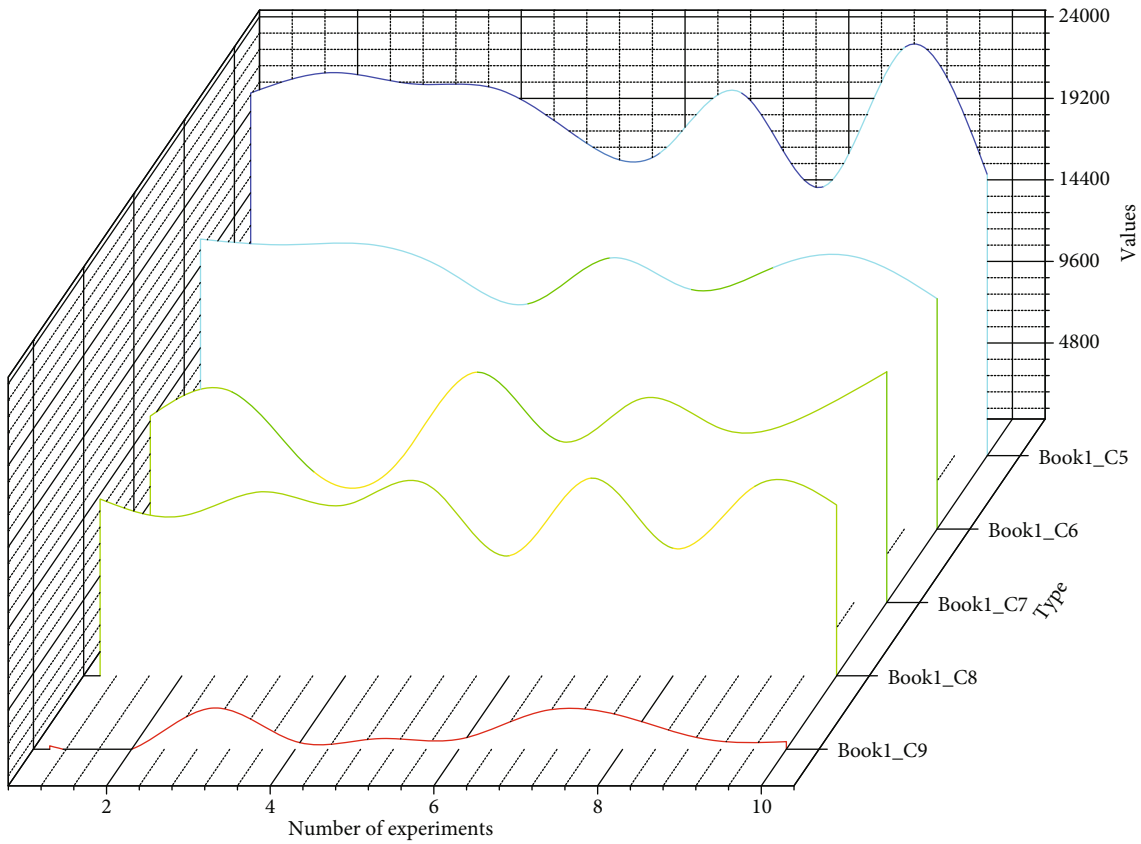


FIGURE 6: Performance comparison of different knowledge point coverage.

coverage of the completed paper is not extensive enough. The automatic paper formation system is based on a randomized algorithm, automatic paper formation system based on a backward matching algorithm, and automatic paper formation system based on a genetic algorithm. Finally, the improved genetic algorithm is implemented on the application of automatic volume formation based on the optimization of coding, cross-variance, and selectors, and the improved genetic algorithm is simulated and analyzed on simulated experimental data. It is necessary to establish an open and relaxed evaluation atmosphere, pay attention to teacher evaluation and encourage students to participate, realize the diversification of evaluation subjects, and help students to reflect on self-evaluation, teacher evaluation, and mutual evaluation.

*4.2. Analysis of English Teaching Reform Results.* The design of the wisdom classroom teaching model in the big data environment provides a reference for the application and innovation of the teaching model, but how to conduct a scientific, comprehensive, and accurate evaluation of the teaching effect in the wisdom classroom teaching has aroused the concern of relevant educators, and the effective customization and optimization of the teaching program by evaluating the application effect of the classroom teaching model are also an urgent research level for the development of information-based education. The quality of “wise” teaching and “intelligent” learning is also in urgent need of research, to improve the quality of “wise” teaching and “wise” learning.

The number of selector selection is 10 times, the number of individuals generated by the crossover operator is 20, the expected test paper difficulty coefficient is 0.67, and the total score for each set of questions is 100 points. The SSA algorithm is used to analyze the data and predict the learning behavior of each student, to detect any risk of misunderstanding during the course. The course instructor can rely on the algorithms and models submitted by the platform to study the learning outcomes and conditions of the students during the teaching process and then intervene to provide the students with the right type of course and resources for their situation. When conducting the analysis, the focus is on obtaining information that includes outcomes and learning behaviors, as well as course performance and learning characteristics, and analyzing students' learning behaviors and performance by providing timely feedback. Interventions are made for poor conditions in them. Through the application of this system, students' performance has all obtained a certain degree of improvement, both in terms of learning effect and learning performance, which obtained a substantial improvement compared to the previous one. It is evident that this intelligent platform is more effective in practical application, and it is based on this effect that other schools have also imitated and adopted this system at the same time, as shown in Figure 7.

From the data analysis of the survey results, we can see that the mean value of students' classroom participation is low except for “I listen attentively, take notes, and pay attention in class.” Students can participate in classroom teaching

activities, listen attentively, take notes, and pay attention in class, but they are not very active in answering questions raised by the teacher, asking questions they do not understand, presenting discussion results, listening to others' reports, and evaluating others, and they cannot actively speak, present reports, and evaluate, and they have not yet developed the habit of active thinking and questioning. They have not yet developed the habit of active thinking and questioning. According to the standard deviation of “I enthusiastically present the results of my group's discussion on the stage in class,” the standard deviation of 1.166 is greater than 1.000, which shows that there is a significant difference in the initiative of different students in the class to present the results of group learning.

As shown in Figure 8, among the students' language, the part of reading aloud accounts for 38%, the proportion of students' passive responses is 33%, the proportion of active responses is 21%, and the proportion of students' active questions is only 1%, so most students have not yet developed the habit of actively responding and actively asking questions. Given this problem, teachers should pay attention to further guiding and encouraging students to think positively, motivating them to express their own opinions, organizing more discussions and exchanges, thinking in discussions, and improving students' enthusiasm to participate in classroom learning. The percentage of students' intergroup assessment is small at 7%, which indicates that students participate in intergroup assessment learning activities less frequently. To address this problem, teachers should further motivate more students to actively participate in intergroup assessment learning activities through certain teaching methods to improve student's evaluation and reflection skills.

Due to the differences in students' abilities and personalities, the presentation session is still the stage for the top students to perform. These students with stronger speaking ability actively speak and perform in the group discussion, help the group to complete the task, and take the initiative to present on the stage. Some students who are introverted and not strong in speaking ability are more passive in the discussion and practice session and have low initiative and participation in the presentation of results. Other students, who are used to the traditional classroom teaching model, are more dependent on the teacher and lack a sense of active learning. For example, students rarely take the initiative to answer questions, think actively in class, and rarely put forward their own opinions or disbeliefs. Teachers should pay more attention to guiding students' thinking by asking questions while reducing learning activities that organize students to read aloud and repeat mechanically and using various forms of interaction to encourage students to answer actively. Some students are introverted and poor in oral English. They are relatively passive in the discussion and practice sessions, have low initiative in the presentation of results, and have low participation. There are also students who are accustomed to the traditional classroom teaching mode and are highly dependent on teachers and lack the awareness of active learning. When introducing new lessons and assigning tasks, with the help

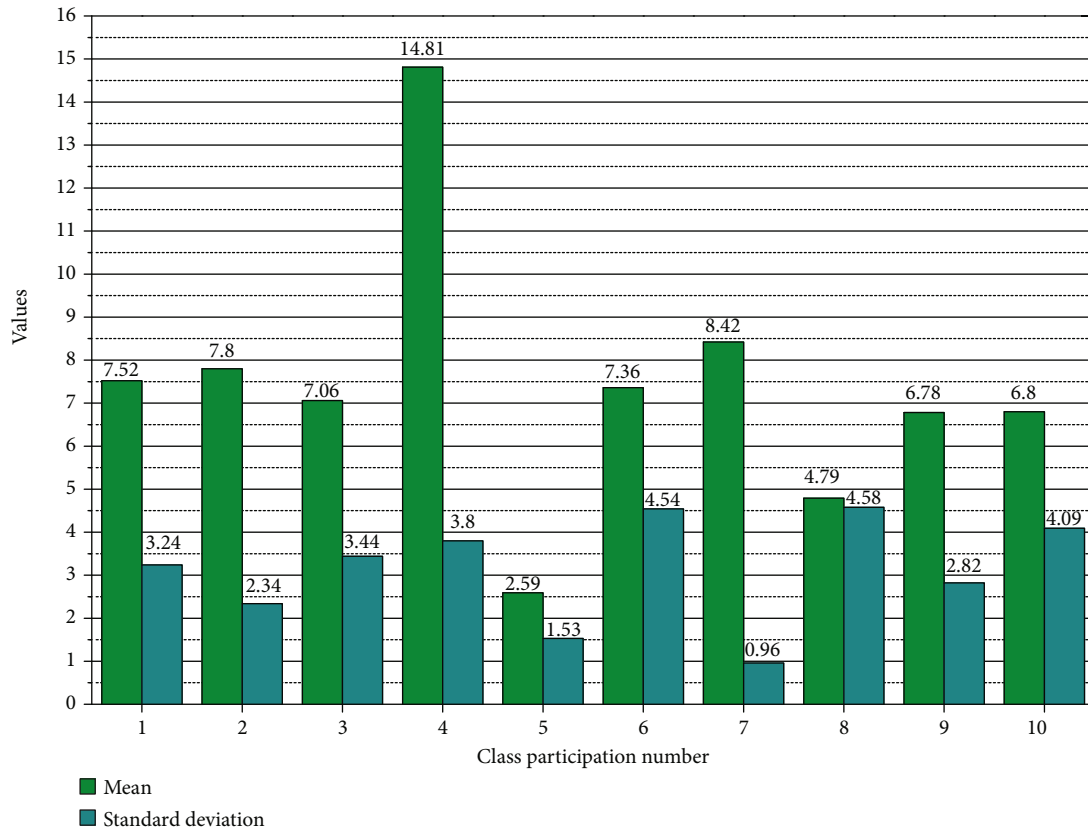


FIGURE 7: Results of the survey on the effectiveness of classroom interaction.

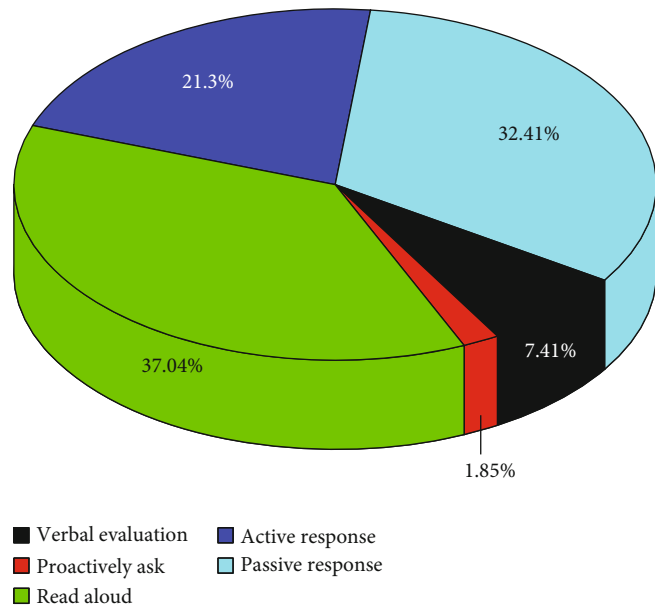


FIGURE 8: Percentage of each part of students' speech.

of interactive whiteboards and tablet PCs and the rich multimedia learning resources in the cloud platform, teachers should introduce topics close to life, create vivid and interesting learning situations to mobilize students' multiple senses, stimulate students' desire to continue exploring

and learning, design questions and tasks that fit the topic and are close to students' nearest developmental zone, with a clear hierarchy, and focus on overall class participation, to effectively promote teacher-student interaction and student-student interaction.



## 5. Conclusion

The fuzzy comprehensive analysis method was used to evaluate and analyze the comprehensive application effect of this model. The factors affecting the teaching effect after the implementation of this model were first investigated and analyzed, and the relevant index sets were designed by adopting the valuable opinions of experts and teachers, and then, the validity of the indexes was verified, and it was concluded that the classroom teaching effect was good after the application of this model. The development and design of the system fundamentally reduce the teachers' workload and improve efficiency. Our existing examination system is to manually correct the papers, which is a very troublesome and heavy workload, so this oral examination system design can change the limitations of this existing oral examination, promote the quality of teaching and learning, and improve students' listening and speaking skills. The system is designed to create a communication platform for teachers, students, and administrators in daily English-speaking exams, simplify the process of speaking exams, make the traditionally tedious and complicated exam process simple and easy to operate, and make speaking exams as easy to operate as other exams. Changing the original one-to-one model saves workforce, strengthens students' training and test frequency, helps improve the efficiency of daily teaching management, and greatly reduces the labor intensity of English teachers. At the same time, the introduction of big data technology improves the efficacy of the English-speaking test system, which in turn improves the efficiency of teachers to focus on teaching and research.

## Data Availability

The data used to support the findings of this study are included within the article.

## Conflicts of Interest

All the authors do not have any possible conflicts of interest.

## Acknowledgments

This study was supported by the following funding: (1) the Key Project of 2020 Educational Science of Shaanxi Province in the 13th Five-Year Plan titled "Research on Classroom Teaching Reform of Colleges and Universities in Shaanxi Province in the New Era" (Grant No. SGH20Z017) and (2) the 2019 Education Reform Project of Shaanxi Province titled "Research and Practice of Smart Classroom Model of College English Based on Dynamic Learning Data Analysis" (Grant No. 19BY131).

## References

- [1] C. Yang, S. Huan, and Y. Yang, "Application of big data technology in blended teaching of college students: a case study on rain classroom," *International Journal of Emerging Technologies in Learning (ijET)*, vol. 15, no. 11, pp. 4–16, 2020.
- [2] C. Meng-yue, L. Dan, and W. Jun, "A study of college English culture intelligence-aided teaching system and teaching pattern," *English Language Teaching*, vol. 13, no. 3, pp. 77–83, 2020.
- [3] C. Fischer, Z. A. Pardos, R. S. Baker et al., "Mining big data in education: affordances and challenges," *Review of Research in Education*, vol. 44, no. 1, pp. 130–160, 2020.
- [4] C. J. Wang, C. Y. Ng, and R. H. Brook, "Response to COVID-19 in Taiwan: big data analytics, new technology, and proactive testing," *JAMA*, vol. 323, no. 14, pp. 1341–1342, 2020.
- [5] C. Liu, Y. Feng, D. Lin, L. Wu, and M. Guo, "Iot based laundry services: an application of big data analytics, intelligent logistics management, and machine learning techniques," *International Journal of Production Research*, vol. 58, no. 17, pp. 5113–5131, 2020.
- [6] L. Liang, Q. Yin, and C. Shi, "Exploring proper names online and its application in English teaching in university," *ASP Transactions on Computers*, vol. 1, no. 1, pp. 24–29, 2021.
- [7] P. Zheng, X. Wang, and J. Li, "Exploration and practice of curriculum ideological and Political Construction Reform —take "information security" course as an example," *ASP Transactions on Computers*, vol. 1, no. 1, pp. 1–5, 2021.
- [8] J. Zhang, Y. Liu, H. Liu, and J. Wang, "Learning local-global multiple correlation filters for robust visual tracking with Kalman filter redetection," *Sensors*, vol. 21, no. 4, p. 1129, 2021.
- [9] W. Cai, Y. Song, and Z. Wei, "Multimodal data guided spatial feature fusion and grouping strategy for E-commerce commodity demand forecasting," *Mobile Information Systems*, vol. 2021, Article ID 5568208, 14 pages, 2021.
- [10] M. Zhao, A. Jha, Q. Liu et al., "Faster mean-shift: GPU-accelerated clustering for cosine embedding-based cell segmentation and tracking," *Medical Image Analysis*, vol. 71, p. 102048, 2021.
- [11] M. M. Hasan, J. Popp, and J. Oláh, "Current landscape and influence of big data on finance," *Journal of Big Data*, vol. 7, no. 1, pp. 1–17, 2020.
- [12] B. Aho and R. Duffield, "Beyond surveillance capitalism: privacy, regulation and big data in Europe and China," *Economy and Society*, vol. 49, no. 2, pp. 187–212, 2020.
- [13] R. Agrawal and S. Prabhakaran, "Big data in digital healthcare: lessons learnt and recommendations for general practice," *Heredity*, vol. 124, no. 4, pp. 525–534, 2020.
- [14] R. Hou, Y. Q. Kong, B. Cai, and H. Liu, "Unstructured big data analysis algorithm and simulation of Internet of Things based on machine learning," *Neural Computing and Applications*, vol. 32, no. 10, pp. 5399–5407, 2020.
- [15] N. C. Burbules, G. Fan, and P. Repp, "Five trends of education and technology in a sustainable future," *Geography and Sustainability*, vol. 1, no. 2, pp. 93–97, 2020.
- [16] G. Boeing, M. Besbris, A. Schachter, and J. Kuk, "Housing search in the age of big data: smarter cities or the same old blind spots?," *Housing Policy Debate*, vol. 31, no. 1, pp. 112–126, 2021.
- [17] J. Knox, "Artificial intelligence and education in China," *Learning, Media and Technology*, vol. 45, no. 3, pp. 298–311, 2020.
- [18] Chengdu Library and Information Center, Chengdu, Sichuan, China, Department of Library, Information and Archives Management, School of Economics and Management, University of Chinese Academy Sciences, Beijing, China, Q. Tian et al., "Research topics and future trends on maker education

- in China based on bibliometric analysis,” *International Journal of Information and Education Technology*, vol. 10, no. 2, pp. 135–139, 2020.
- [19] N. Zhao, X. Zhou, B. Liu, and W. Liu, “Guiding teaching strategies with the education platform during the COVID-19 epidemic: taking Guiyang No. 1 Middle School teaching practice as an example,” *Science Insights Education Frontiers*, vol. 5, no. 2, pp. 531–539, 2020.
- [20] Z. Yu, “Visualizing artificial intelligence used in education over two decades,” *Journal of Information Technology Research*, vol. 13, no. 4, pp. 32–46, 2020.
- [21] M. Xiao and H. Yi, “Building an efficient artificial intelligence model for personalized training in colleges and universities,” *Computer Applications in Engineering Education*, vol. 29, no. 2, pp. 350–358, 2021.
- [22] K. Yang, “The construction of sports culture industry growth forecast model based on big data,” *Personal and Ubiquitous Computing*, vol. 24, no. 1, pp. 5–17, 2020.
- [23] Jiangsu Second Normal University, Nanjing 211200, Jiangsu, China, L. Zhou, Engineering Research Center of Digital Learning Support Technology, Ministry of Education, Changchun 130000, Jilin, China et al., ““School’s Out, But Class’s On”, the largest online education in the world today: taking China’s practical exploration during the COVID-19 epidemic prevention and control as an example,” *Best Evidence of Chinese Education*, vol. 4, no. 2, pp. 501–519, 2020.
- [24] G. J. Hwang and Q. K. Fu, “Advancement and research trends of smart learning environments in the mobile era,” *International Journal of Mobile Learning and Organisation*, vol. 14, no. 1, pp. 114–129, 2020.
- [25] M. Gao, W. Cai, and R. Liu, “AGTH-net: attention-based graph convolution-guided third-order hourglass network for sports video classification,” *Journal of Healthcare Engineering*, vol. 2021, Article ID 8517161, 10 pages, 2021.
- [26] Y. Gu, A. Chen, X. Zhang, C. Fan, K. Li, and J. Shen, “Deep learning based cell classification in imaging flow cytometer,” *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 2, pp. 18–27, 2021.
- [27] M. N. Habib, W. Jamal, U. Khalil, and Z. Khan, “Transforming universities in interactive digital platform: case of city university of science and information technology,” *Education and Information Technologies*, vol. 26, no. 1, pp. 517–541, 2021.
- [28] J. Wan, J. Li, Q. Hua, A. Celesti, and Z. Wang, “Intelligent equipment design assisted by Cognitive Internet of Things and industrial big data,” *Neural Computing and Applications*, vol. 32, no. 9, pp. 4463–4472, 2020.
- [29] P. Layton, “Artificial intelligence, big data and autonomous systems along the belt and road: towards private security companies with Chinese characteristics?,” *Small Wars & Insurgencies*, vol. 31, no. 4, pp. 874–897, 2020.
- [30] S. Lv, “Construction of marine ship automatic identification system data mining platform based on big data,” *Journal of Intelligent & Fuzzy Systems*, vol. 38, no. 2, pp. 1249–1255, 2020.

## Research Article

# Intelligent Learning Algorithm for English Flipped Classroom Based on Recurrent Neural Network

Qi Shan 

*Jiangsu Union Technical Institute, Xuzhou Jiangsu 221000, China*

Correspondence should be addressed to Qi Shan; shanqi60032@163.com

Received 12 July 2021; Revised 8 August 2021; Accepted 17 August 2021; Published 13 September 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Qi Shan. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Reading and writing are the foundations of English learning as well as an important method of instruction. With the advancement of network technology and the onset of the information age, an increasing number of students have lost interest in traditional English reading and writing instruction in the classroom. Flipped classrooms have emerged as a result of this situation and have become the focus of research in one fell swoop. As a result, flipped classroom research at home and abroad has primarily focused on the theory and practical application of flipped classrooms, and flipped classroom application practice is primarily based on the overall classroom, with few separate discussions on the effects of flipped classroom students' self-learning. As a result, we developed a recurrent neural network-based intelligent assisted learning algorithm for English flipped classrooms. There are two main characteristics of the model. First, it is a gated recurrent unit based on a variant structure of the recurrent neural network. The double-gating mechanism fully considers the context and selects memory through weight assignment, and on this basis, it integrates the novel LeakyReLU function to improve the model's training convergence efficiency. Second, by overcoming time-consuming problems in the medium, the adoption of the connection sequence classification algorithm eliminates the need for prior alignment of speech and text data, resulting in a direct boost in model training speed. The experimental results show that in the English flipped classroom's intelligent learning mode, students explore and discover knowledge independently, their enthusiasm and interest in learning are greatly increased, and the flipped classroom's teaching effect is greatly improved.

## 1. Introduction

Education experts advocate a new teaching model in the context of the new curriculum reform [1–3], in which students are the main body and teachers are the leaders. This concept, however, cannot be fully reflected in classroom teaching in specific practical situations. The flipped classroom [4–6] came into being under this situation, which is in line with the requirements of contemporary educational concepts. Flipped classroom advocates autonomous learning. In flipped classroom teaching, teachers should not only teach students professional knowledge but also provide personalized guidance in the classroom to teach students the ability to learn. This is not only the requirement of the times but also the direction of the future development of national education.

Secondly, the current situation of literacy class teaching is unsatisfactory, which leads to worrying students' literacy skills [7, 8]. First, in the process of English reading and writing teaching [9], teachers and students lack sufficient attention to reading ability, writing ability, and communication ability. Second, the atmosphere in the teaching of English reading and writing is tense and fast. The English books that students can access every day are extremely limited. In addition to English textbooks, there are volumes of exercises and sets of exercises. As a result, the current stage of English reading and writing course teaching content has significant limitations and has failed to effectively and successfully broaden the scope of knowledge [10]. Many English reading and writing teachers do not know how to organize classroom activities, and some teachers do know how to organize classroom activities, but the effect is clearly difficult to

achieve expectations. To carry out classroom activities in English reading and writing classes, it is necessary to combine the flipped classroom teaching mode [11–13].

The ability of reading and writing is one of the main abilities of English learning, which occupies a vital position. For many years, the traditional teaching of English reading and writing has been based on students' passive acceptance, which has led to many students' boredom of reading and writing, resulting in the decline of students' learning confidence [5, 14], which makes it difficult for students to improve their English scores. Many scholars are studying how to change this situation. In English reading and writing teaching, using modern information technology as an auxiliary teaching means, improve students' learning confidence and interest, via class activities, read presses to write, to write gong of reading, speaking, reading, and writing as a whole, guide students to learn actively.

However, the speech signal of English reading aloud [15] has its complexity. It needs to take into account the students' different accents, volume, speed, pauses, and other factors. The input data has specificity due to different users, making the input of speech data more complicated and more complicated. Diversification adds a lot of difficulty to the recognition and training process of the model. According to the different types of speech data processed, it can be divided into isolated word recognition [16, 17], small vocabulary continuous speech recognition [18, 19], and large vocabulary continuous speech recognition [20, 21]. This article mainly focuses on the research of medium and long sentences with large vocabulary. In the structure of long sentences with large vocabulary, the context contains greater relevance. This relevance has a great effect on the learning process of the model. Many training models often ignore such relevance and only learn the current state. A single mapping relationship leads to low efficiency. Combining the characteristics of recurrent neural networks in machine learning [22], this article believes that assigning a certain "memory" weight to the relevant information of the context will help the judgment of the current output and more accurately convert the time sequence information of speech into the corresponding correct characters. This article will focus on the research and analysis of the variant structure GRU under the recurrent neural network branch [23, 24], make some improvements to its internal calculation structure, and combine it with a popular time series classification method [25, 26] to reduce the error rate in recognition results.

The main innovations of this paper are as follows:

- (1) I built an intelligent assisted learning algorithm for English flipped classroom based on recurrent neural network, and the experimental results proved that in the intelligent learning mode of English flipped classroom, students autonomously explore and discover knowledge, and their learning enthusiasm and interest are greatly enhanced, flipping the classroom. The teaching effect has also been greatly improved

- (2) To solve the existing problems in the field of English speech recognition, a hybrid cyclic neural network model incorporating LeakyReLU functions is proposed in this paper. The characteristics of the model are mainly based on a variant structure gated loop unit of the neural network. The relevance of the context is fully considered through the double-gated mechanism, selective memory is carried out through weight assignment, and the joining time sequence classification algorithm is introduced to improve the speed of model training

The remainder of this paper is organized as follows: Section 2 analyzes some related work. Some details of the principles of the proposed algorithm and related submodules are introduced in Section 3. Section 4 provides details of experimental results. In Section 5, the conclusion based on this study is given.

## 2. Related Work

Robert modeled the flipped classroom teaching process and clearly divided flipped classrooms. Specifically, it can be divided into two stages: before class and during class. The preclass mainly includes students watching instructional videos for target teaching and completing well-designed preclass exercises [27–29]; the inclass stage includes a small amount of teacher evaluation, student problem-solving, problem feedback, and evaluation of four steps. Through the analysis of the current popular flipped classroom teaching model, it can be found that most of the flipped classroom teaching process models built at home and abroad are based on Robert's structural model diagram as the main prototype and then combine different learning theories to target specific teaching situations. Make changes and start research. Aaron Sams believes that teachers can not only use the excellent teaching videos of other teachers but also use the Camtasia Studio screen recording software and other screen recording tools to make their own teaching videos, but they should ensure that the videos are short, vivid, and clear. With the deepening of flipped classroom research and practice, educators believe that to a large extent, the learning effect of students' preclass teaching videos determines the effect of classroom teaching. As a result, more focus is placed on assessment, as well as teachers and students, in order to communicate with preclass video learning impacts. Chung et al. used the "first guiding principle" of meta-design theory to design a flipped classroom method and conducted a two-stage study in two middle schools. Participants had a total of 382 students and 5 teachers, who came from four subject areas, namely, mathematics, physics, Chinese, and information and communication technology (ICT). Based on the experience of the pilot study, the author improved the flipped classroom model and tested its effect through the quasiexperimental design in the main study. The results found that the learning outcomes of the students in the flipped ICT course were similar to those of the nonflip ICT courses, but the student performance levels of the other three courses (namely,

mathematics, physics, and Chinese) improved after the flip, with small- and medium-sized effects. Finally, some practical suggestions are put forward in the article, and ideas for further research are put forward.

In addition, Sarah et al. look at effective classroom activities in English as a foreign language (EFL) and propose a student response system (SRS) to support teachers in organizing classroom activities in flipped classes. In order to investigate the effectiveness of this method, a quasiexperiment was conducted in an EFL classroom of an engineering school. The experimental group used SRS for classroom activities, while the control group followed the usual method. The results show that the use of SRS can improve students' motivation and self-efficacy in learning English grammar and improve their participation in classroom activities during flipped learning. In addition, the results of the questionnaire survey show that students accept SRS as a teaching method in EFL flipped classroom. Hernan et al. [30] described a teaching intervention that included a range of creative activities aimed at improving students' oral and writing English, particularly those who showed interest or attention deficits. At first, the participants did not appear to be particularly interested in learning the language. In the end, they were more willing and motivated to participate in chain games, creative writing, and screenwriting exercises after hearing about the proposed approach. These activities have aided students in improving their spoken and written English fluency as well as their comprehension of English grammar and structure.

### 3. Methodology

*3.1. Definition of Flipped Classroom.* The so-called flipped classroom is relative to the current teaching method of teacher explanations, students listening to lectures, and student homework after class; it refers to the use of information technology to facilitate teachers to record the explanations of knowledge points into short and succinct teaching methods. The video, accompanied by other learning materials and advanced homework, is sent to students through the learning management platform. Under the guidance of the teacher, the students will conduct self-study and complete the advanced homework; based on the information on the learning platform, the teacher is grasping the learning situation in detail. Under the circumstance, the classroom will focus on the targeted explanation, solve the problems with the students, and complete the homework. In the flipped classroom teaching model, the acquisition of knowledge mainly relies on students' self-study before class and no longer takes up students' valuable classroom time for teachers to teach. Students can use a variety of channels: online discussions with other students, watching related video materials, reading required image materials, etc. In the time in the classroom, students can learn more actively based on the information obtained before class to gain a deeper understanding. Teachers can also make full use of the remaining spare time to communicate with everyone. After class teaching, students plan their own learning content, learning pace, and learning style and adopt appropriate

knowledge presentation methods; teachers use cooperation and teaching methods to meet the needs of students, promote students' personalized inquiry learning, and let students get a real learning experience in self-inquiry.

The essence of the flipped classroom is to return the dominance of learning to the students, lead the students' subjective teaching methods, and have the energy to create education for the future. And flipped classroom as a kind of teaching concept and teaching mode is influencing and changing the traditional classroom teaching. It makes use of Internet technology and information technology to break down traditional classroom barriers, expand classroom teaching time and space, optimize students' learning processes, improve students' learning abilities, and realize in-depth integration of information technology and curriculum teaching, all of which promote students' deep learning. At the same time, the flipped classroom model, as a part of the education reform movement, will completely subvert the traditional printing-based classroom teaching structure and teaching process and trigger a series of changes in the role of teachers, curriculum models, and management models.

Autonomous learning, in contrast to traditional acceptance learning, places a greater emphasis on students' ability to learn independently. As the subject of learning, students achieve their own learning goals through independent analysis, exploration, practice, and creation. We should change the current situation of curriculum implementation that emphasizes too much on learning, rote memorization, and mechanical learning, advocate students' active participation, willing to explore and diligent in doing, and cultivate students' ability to collect and process information, acquire new knowledge, analyze and solve problems, and communicate and cooperate. The concept of autonomous learning is based on inquiry-based learning, in which students are presented with situations in which they must conduct their own research, solve problems, and gain knowledge in the field. As a form of discovery learning, students in the classroom are provided with more opportunities to "experience and interact" with knowledge. It is promising to use the current popular neural network technology [31–33] to assist English learning.

#### 3.2. English Speech Recognition

*3.2.1. Language Model.* There are often multiple combinations of output sequences that can meet the original signal input conditions in the speech recognition process. An external tool is required to assist the machine in constraining the discriminative category of words when judging the output, resulting in a more logical collocation situation. The role of the language model is to summarize the most probable text sequence on the output of the acoustic model, just like a dictionary. In the method of processing continuous speech recognition with large vocabulary, the commonly used language model is the  $N$ -gram language model, which is characterized by using the relationship between  $N$  adjacent words in the context to calculate the current maximum probability output. The  $N$ -gram model

is based on the assumption that the current output is only related to the previous  $N - 1$  words and has no relationship with other words. Assuming a word sequence is  $W = (W_1, W_2, \dots, W_N)$ , then its probability calculation process is as follows:

$$P(W) = P(W_1, W_2, \dots, W_N) = \prod_{i=1}^N P(W_i | W_{i-N+1}, \dots, W_{i-1}), \quad (1)$$

where  $N$  is the number of words in the sequence and  $W_i$  is the  $i$ -th word in the sequence  $W$ .

There are certain restrictions on the selection of the  $N$  value in  $N$ -gram. Generally, 1~5 is more appropriate. If the value is too large, it will cause data sparseness. The more commonly used are binary language model (bi-gram) and the ternary language model (tri-gram). Among them, bi-gram only considers the impact of its previous word on the current judgment, and its probability calculation equation is as follows:

$$P(W) = P(W_1, W_2, \dots, W_N) = \prod_{i=1}^N P(W_i | W_{i-1}), \quad (2)$$

where  $P(W_i | W_{i-1})$  can be calculated by maximum likelihood estimation:

$$P(W_i | W_{i-1}) = \frac{\text{count}(W_{i-1}W_i)}{\text{count}(W_{i-1})}, \quad (3)$$

where  $\text{count}(W_{i-1}W_i)$  is the frequency of occurrence of the phrase  $W_{i-1}W_i$  in the transcription text set and  $\text{count}(W_{i-1})$  is the frequency of occurrence of the word  $W_{i-1}$ . Similarly, tri-gram considers the influence of the first two words on the current judgment.

**3.3.2. Pronunciation Dictionary.** The pronunciation dictionary keeps track of the various characters to which each pronunciation can respond. It is generally identified by various linguistics official records and can be used directly in recognition methods to aid learning. In Chinese, the corresponding relationship is pinyin and Chinese characters. For example, when learning a Chinese character, we know how to pronounce it, but we do not know which character it corresponds to. We need to use tools like ‘‘Xinhua Dictionary’’ to assist in learning. Similarly, in the process of voice recognition, it also needs the participation of such a tool. Following the previous example, in the pinyin part, ‘‘ta’’ in the pronunciation dictionary can correspond to the Chinese characters ‘‘it,’’ ‘‘he,’’ ‘‘she,’’ etc., which compresses the recognition results within certain constraints. The model is then used to match the most likely characters using the process of learning context information. The function of the pronunciation dictionary, from this perspective, is to match the phoneme information from the acoustic model with the word information from the language model. It is a highly effective tool for acoustic recognition.

### 3.3. Our Model

**3.3.1. Gated Recurrent Unit.** The intuitive improvement of gated recurrent unit is to merge the three gate units in the LSTM network into two gate units, which is more concise and refined in structure and is a good deformed structure. As a modified structure of LSTM, LSTM can solve the long dependency problem in RNN, as can GRU. On this basis, GRU also reduces the number of parameters, avoiding the problem of slower training speed due to more and more parameters in the longer-term training and learning of LSTM.

Figure 1 is the unit state of the previous unit,  $h_{t-1}$  is the input of the current unit, and they are collectively used as the calculation input of the current unit, and  $h_{t-1}$  carries the information passed by the previous unit selection. GRU turns the input gate, forget gate, and output gate in LSTM into two gates: update gate  $z_t$  and reset gate  $r_t$ . Among them, the update gate is mainly responsible for judging the degree to which the transfer information of the previous unit at the previous moment is calculated into the current state. The larger the value of the update gate, the deeper the state information previously transferred is calculated into the current state. The reset gate is mainly responsible for controlling how much information of the previous unit state is written to the candidate set of the current unit. The smaller the reset gate, the less information of the previous state is written. Finally, the unit state and output are combined into a state  $h_t$ , which makes the model have fewer parameters and higher processing efficiency.

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z), \quad (4)$$

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r), \quad (5)$$

$$h'_t = \tanh(W_h x_t + U_h (h_{t-1} r_t) + b_h), \quad (6)$$

$$h_t = z_t h_{t-1} + (1 - z_t) h'_t. \quad (7)$$

The place where GRU is more efficient than LSTM is that it uses one step to complete the two actions of selective memory and selective forgetting.

**3.3.2. Connectionist Temporal Classification.** The connectionist temporal classification (CTC) was first proposed by Graves et al., as a temporal classification algorithm. Mainly used for the alignment of speech signals and text labels, CTC can help save the time cost of human alignment and improve training efficiency. Traditionally used cross entropy loss requires that the tags of the training data be aligned at the frame level before processing the speech sequence data. The alignment operation requires a certain amount of work, and the model needs to know the corresponding tags of each frame before training, which reduces work efficiency. In the process of training, CTC can automatically optimize the mapping relationship between the input sequence and the output sequence for direct training, which greatly improves the speed of decoding.

We use CTC to process the GRU output information. The output length of GRU is the same as the input length.

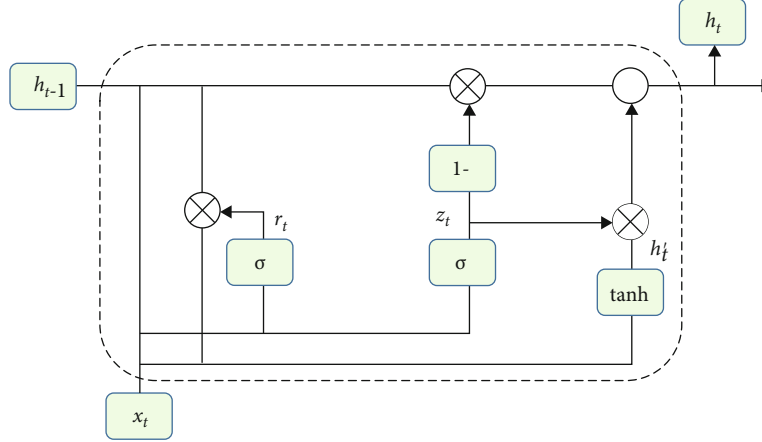


FIGURE 1: Schematic diagram of GRU unit structure.

With  $N$  voice signal inputs, there are  $N$  probability vector outputs, including the probability of each character corresponding to the dictionary. When the voice signal  $X$  is input,

$$\bar{Y} = \underset{Y}{\operatorname{argmax}} P(Y | X), \quad (8)$$

where  $\bar{Y}$  is the output with the highest conditional probability and  $P(Y|X)$  is the posterior probability of  $Y$ . During training, I hope to obtain its maximum value, that is, the most likely output. In recognition processing, the nearest similar output can be found more quickly.

As shown in Figures 2 and 3, the CTC algorithm does not require the input and output to be strictly aligned. Multiple output paths may correspond to one output result. Understanding the corresponding relationship between input and output can help us better understand the calculation method of loss function and the calculation method used in the test. The most prominent feature of CTC is the introduction of a placeholder blank node, mainly to model the parts without effective information, such as silence and pause, to represent the output state of the network when predicting uncertain information. After the alignment of the blank, it needs to be deleted. This process is called  $F$  transformation. If an output sequence of the network can be mapped to the correct annotated sequence through  $F$  transformation, then the output sequence is a CTC path. The process of  $F$  transformation is as follows: first, delete the repeated annotations between adjacent blank nodes in the sequence and then delete the blank nodes.

**3.3.3. LeakyReLU Improved GRU.** The calculation equation of LeakyReLU function is as follows:

$$\operatorname{LeakyReLU}(x) = \begin{cases} x, & x \geq 0, \\ ax, & x < 0, \end{cases} \quad (9)$$

where  $a$  is the fixed value of the slope selected before training, with the purpose of making the nonpositive part have a small slope. Compared with the training characteristics of sigmoid and tanh function, it is more consistent with

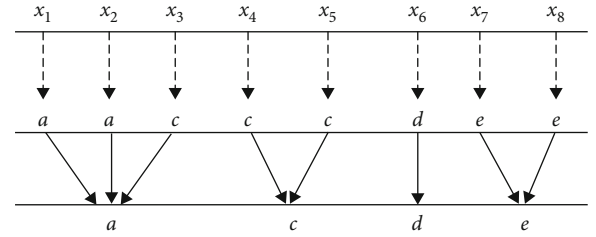


FIGURE 2: Schematic diagram of traditional alignment.

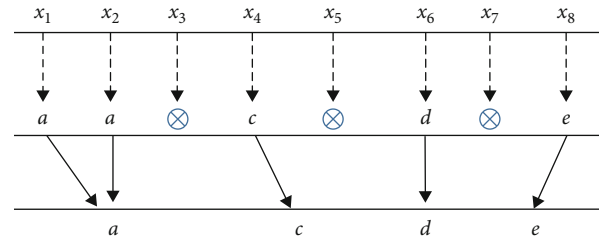


FIGURE 3: Schematic diagram of the alignment process of CTC placeholders.

the convergence effect that the training wants to achieve, so that the training process of the model can be stable and efficient.

## 4. Experiments

**4.1. Parameter Settings.** The experiments with all the algorithms were performed on a computer equipped with a single NVIDIA GTX1080TI GPU (11 GB). We have implemented the model construction through the PyTorch deep learning library, the programming language we use is Python, and we batch processed 100 samples each time.

**4.2. Datasets.** The THCHS30 database can be used to build a comprehensive Chinese speech recognition benchmark system, including performance under high noise conditions. The voice data from the dataset was recorded by 40 people in a quiet office environment through a carbon particle microphone. The total time is 35 hours. Most of those

involved in the recording were college students who spoke Mandarin. The recorded content text is from large-capacity news, the sampling frequency is 16kHz, the sampling size is 16 bits, and the format is WAV. The composition of the dataset is shown in Table 1.

**4.3. Evaluation Methods.** The evaluation methods I use are character error rate, sentence error rate, and word error rate. The calculation equations are as follows:

$$\text{CER} = \frac{S_C + D_C + I_C}{N_C} \times 100\%, \quad (10)$$

$$\text{SER} = \frac{S_W + D_W + I_W}{N_W} \times 100\%, \quad (11)$$

where  $S_C$  is the number of replacement characters,  $D_C$  is the number of deleted characters,  $I_C$  is the number of inserted characters, and  $N_C$  is the number of characters.  $S_W$  is the number of replacement words,  $D_W$  is the number of deleted words,  $I_W$  is the number of inserted words, and  $N_W$  is the number of words.

**4.4. Experimental Results.** To test the sexual superiority of GRU in relation to phonological recognition patterns in other languages, LSTM-HMM, LSTM-CTC, and GRU-CTC were selected as the models for the comparative experiment. Through the comparison of CER and SER of the processing results of the same data set of each model in the experimental environment, the smaller the value, the better the performance in the experiment.

From the data in Table 2, the recognition effect of the hybrid model with CTC is lower than the CER and SER of the traditional HMM hybrid model. This set of comparison results shows that the introduction of the CTC method can help the performance of the model to a certain extent. The recognition effect has been improved to a certain extent, mainly due to the fact that the CTC method can directly model the entire sentence when the input and output are not aligned, which improves the completeness and relevance of the entire sentence recognition. The CER and SER of the GRU-CTC model, like the previous control group LSTM-CTC, combined with the CTC approach, are the best among the three deep learning-based models selected in the experiment and LSTM. CER and SER are both reduced as compared to CTC. Combining the above results and the characteristics of CTC and GRU, it can be concluded that CTC, as an efficient method that does not require manual label alignment, and the combination of LSTM, GRU, and other cyclic neural network variants have achieved considerable results. The improved double-door variant structure is more reasonable in the deletion and optimization of the structure, and the control of the relevance of the sequence context information is more efficient, and the recognition performance of the model has been improved. In summary, the hybrid cyclic neural model of GRU and CTC is a relatively efficient method of speech recognition.

**4.5. Ablation Experiment of Activation Function.** In order to further verify the role played by the activation function in

TABLE 1: Components of THCHS30 dataset.

Features	Train	Test
Audio duration	27.23	6.25
Number of speakers	30	10
Male	8	1
Female	22	9
Age	20-55	19-50
Total words	215995	49085

TABLE 2: Comparative experiment results.

Methods	Train (CER)	Train (SER)	Test (CER)	Test (SER)
LSTM-HMM	0.3525	0.4352	0.2874	0.3325
LSTM-CTC	0.3415	0.3952	0.2956	0.3425
GRU+CTC (ours)	0.3325	0.3699	0.2811	0.3155

TABLE 3: Ablation experiment results.

Methods	Train (CER)	Train (SER)	Test (CER)	Test (SER)
ELU	0.3625	0.4158	0.2858	0.3214
ReLU	0.3398	0.4012	0.2915	0.3396
LeakyReLU (ours)	0.3325	0.3699	0.2811	0.3155

the proposed algorithm, we added an ablation experiment. We selected LeakyReLU, ReLU, and ELU, which are the three activation functions for experiments. The experimental results are shown in Table 3.

It can be clearly seen from Table 3 that after using the three activation functions, respectively, the performance of LeakyReLU is the best, and secondly, the performance of ReLU is better than that of ELU. Therefore, it is reasonable and effective to choose the LeakyReLU activation function in this paper.

## 5. Conclusion

In this paper, we propose a recurrent neural network-based intelligent assisted learning algorithm for creating an English flipped classroom. There are two main characteristics of this model. For starters, it is a gated recurrent unit based on a recurrent neural network's variant structure. On this basis, the dual-gating mechanism fully considers the context's relevance, selects memory through weight allocation, and integrates the novel LeakyReLU function, which can improve the model's training convergence efficiency. Second, the connection sequence classification algorithm can be used to solve the general model. The time-consuming problem of processing input and output alignment is moderate, and model training speed has increased. The experimental results show that in the English flipped classroom's intelligent learning mode, students explore and discover



knowledge independently, and their enthusiasm and interest in learning are greatly enhanced, as is the flipped classroom's teaching effect.

## Data Availability

The data used to support the findings of this study are included within the article.

## Conflicts of Interest

The author does not have any possible conflicts of interest.

## References

- [1] P. Zheng, X. Wang, and J. Li, "Exploration and practice of curriculum ideological and political construction reform—take "information security" course as an example," *ASP Transactions on Computers*, vol. 1, no. 1, pp. 1–5, 2021.
- [2] F. Moller and T. Crick, "A university-based model for supporting computer science curriculum reform," *Journal of Computers in Education*, vol. 5, no. 4, pp. 415–434, 2018.
- [3] K. Pyhältö, J. Pietarinen, and T. Soini, "Dynamic and shared sense-making in large-scale curriculum reform in school districts," *The Curriculum Journal*, vol. 29, no. 2, pp. 181–200, 2018.
- [4] G. Akçayır and M. Akçayır, "The flipped classroom: a review of its advantages and challenges," *Computers & Education*, vol. 126, pp. 334–345, 2018.
- [5] I. T. Awidi and M. Paynter, "The impact of a flipped classroom approach on student learning experience," *Computers & Education*, vol. 128, pp. 269–283, 2019.
- [6] P. Strelan, A. Osborn, and E. Palmer, "The flipped classroom: a meta-analysis of effects on student performance across disciplines and education levels," *Educational Research Review*, vol. 30, p. 100314, 2020.
- [7] N. Amalia, *Narrative Writing Intervention Plan: Analysis of Students' Literacy Learning Needs*, Muhammadiyah University Press, Indonesia, 2017.
- [8] J. Castellví, M. C. Díez-Bedmar, and A. Santisteban, "Pre-service teachers' critical digital literacy skills and attitudes to address social problems," *Social Sciences*, vol. 9, no. 8, p. 134, 2020.
- [9] L. Liang, Q. Yin, and C. Shi, "Exploring proper names online and its application in English teaching in university," *ASP Transactions on Computers*, vol. 1, no. 1, pp. 24–29, 2021.
- [10] A. R. Kothari, J. J. Bickford, N. Edwards, M. J. Dobbins, and M. Meyer, "Uncovering tacit knowledge: a pilot study to broaden the concept of knowledge in knowledge translation," *BMC Health Services Research*, vol. 11, no. 1, pp. 1–10, 2011.
- [11] Z. Turan and B. Akdag-Cimen, "Flipped classroom in English language teaching: a systematic review," *Computer Assisted Language Learning*, vol. 33, no. 5–6, pp. 590–606, 2020.
- [12] M. Y. Abdullah, S. Hussin, and K. Ismail, "Implementation of flipped classroom model and its effectiveness on English speaking performance," *International Journal of Emerging Technologies in Learning*, vol. 14, no. 9, 2019.
- [13] B. Ayçiçek and T. Y. Yelken, "The effect of flipped classroom model on students' classroom engagement in teaching English," *International Journal of Instruction*, vol. 11, no. 2, pp. 385–398, 2018.
- [14] S. Karabatak and H. Polat, "The effects of the flipped classroom model designed according to the ARCS motivation strategies on the students' motivation and academic achievement levels," in *Education and Information Technologies*, pp. 1–21, Springer, 2019.
- [15] E. P. Johnson, B. F. Pennington, J. H. Lowenstein, and S. Nittrouer, "Sensitivity to structure in the speech signal by children with speech sound disorder and reading disability," *Journal of Communication Disorders*, vol. 44, no. 3, pp. 294–314, 2011.
- [16] T. Dijkstra, A. Wahl, F. Buytenhuijs et al., "Multilink: a computational model for bilingual word recognition and word translation," *Bilingualism: Language and Cognition*, vol. 22, no. 4, pp. 657–679, 2019.
- [17] P. Karageorgos, T. Richter, M. B. Haffmans, J. Schindler, and J. Naumann, "The role of word-recognition accuracy in the development of word-recognition speed and reading comprehension in primary school: a longitudinal examination," *Cognitive Development*, vol. 56, p. 100949, 2020.
- [18] V. Mitra, G. Sivaraman, C. Bartels et al., "Joint modeling of articulatory and acoustic spaces for continuous speech recognition tasks," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5205–5209, New Orleans, LA, USA, 2017, March.
- [19] N. H. Gebreegziabher and A. Nürnberger, "An Amharic syllable-based speech corpus for continuous speech recognition," in *In International Conference on Statistical Language and Speech Processing*, pp. 177–187, Springer, Cham, 2019.
- [20] K. Audhkhasi, B. Kingsbury, B. Ramabhadran, G. Saon, and M. Picheny, "Building competitive direct acoustics-to-word models for English conversational speech recognition," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4759–4763, Calgary, AB, Canada, 2018, April.
- [21] Z. Zeng, Y. Khassanov, V. T. Pham, H. Xu, E. S. Chng, and H. Li, "On the end-to-end solution to Mandarin-English code-switching speech recognition," 2018, <http://arxiv.org/abs/1811.00241>.
- [22] W. Chu, P. S. Ho, and W. Li, "An adaptive machine learning method based on finite element analysis for ultra low-k chip package design," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, pp. 1–1, 2021.
- [23] R. Parthiban, R. Ezhilarasi, and D. Saravanan, "Optical character recognition for English handwritten text using recurrent neural network," in *In 2020 International Conference on System, Computation, Automation and Networking (ICSCAN)*, pp. 1–5, Pondicherry, India, 2020, July.
- [24] J. Wang, "Speech recognition in English cultural promotion via recurrent neural network," *Personal and Ubiquitous Computing*, vol. 24, no. 2, pp. 237–246, 2020.
- [25] J. Zhang, W. Wang, C. Lu, J. Wang, and A. K. Sangaiah, "Light-weight deep network for traffic sign classification," *Annals of Telecommunications*, vol. 75, no. 7–8, pp. 369–379, 2020.
- [26] J. Zhang, J. Sun, J. Wang, and X. G. Yue, "Visual object tracking based on residual network and cascaded correlation filters," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 8, pp. 8427–8440, 2020.
- [27] A. T. Taylor, E. L. Olofson, and W. R. Novak, "Enhancing student retention of prerequisite knowledge through pre-class activities and in-class reinforcement," *Biochemistry and Molecular Biology Education*, vol. 45, no. 2, pp. 97–104, 2017.

- [28] J. Jovanovic, N. Mirriahi, D. Gašević, S. Dawson, and A. Pardo, "Predictive power of regularity of pre-class activities in a flipped classroom," *Computers & Education*, vol. 134, pp. 156–168, 2019.
- [29] E. Han and K. C. Klein, "Pre-class learning methods for flipped classrooms," *American Journal of Pharmaceutical Education*, vol. 83, no. 1, p. 6922, 2019.
- [30] C. J. Hernan, T. A. Collins, J. Q. Morrison, and S. D. Kroeger, "Decreasing inappropriate use of mobile devices in urban high school classrooms: comparing an antecedent intervention with and without the good behavior game," *Behavior Modification*, vol. 43, no. 3, pp. 439–463, 2019.
- [31] Y. Ding, X. Zhao, Z. Zhang, W. Cai, and N. Yang, "Multiscale graph sample and aggregate network with context-aware learning for hyperspectral image classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 4561–4572, 2021.
- [32] Y. Tong, L. Yu, S. Li, J. Liu, H. Qin, and W. Li, "Polynomial fitting algorithm based on neural network," *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 1, pp. 32–39, 2021.
- [33] Z. Chu, M. Hu, and X. Chen, "Robotic grasp detection using a novel two-stage approach," *ASP Transactions on Internet of Things*, vol. 1, no. 1, pp. 19–29, 2021.

## Research Article

# Research on Flipped Classroom of Big Data Course Based on Graphic Design MOOC

**Yanqi Wang** 

*Kunming of Department of Fine Art and Art Design University, Kunming 650214, China*

Correspondence should be addressed to Yanqi Wang; [t17301099@stu.ahu.edu.cn](mailto:t17301099@stu.ahu.edu.cn)

Received 23 July 2021; Revised 11 August 2021; Accepted 18 August 2021; Published 9 September 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Yanqi Wang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the rapid development of the Internet, traditional teaching models can no longer meet the needs of talent training in colleges and universities, and reform is imperative. With the advent of the era of big data, the emergence of a large number of rich and diverse teaching resources, MOOC (Massive Online Open Course), microclasses, flipped classrooms, and other teaching models on the Internet has provided reform thinking and directions for teaching reform. This model divides the entire teaching design into two major modules: SPOC (Small Private Online Course) platform teaching activity design and flipped classroom teaching activity design, and applies this model to the actual teaching of open education, designing detailed teaching activity plans, in a real teaching situation. This study uses questionnaire surveys and interview surveys to investigate the basic personal situation of course learners, learning expectations, course participation, learning experience, and learning effects. It is planned to use the questionnaire star platform to issue and return questionnaires and use EXCEL and SPSS software to analyze the data and perform analysis and processing, combined with in-depth interviews with learners and professors for comprehensive analysis, so as to obtain the most true views of students and teachers on this model. In this process, we collect a variety of data from the SPOC platform and the flipped classroom platform, including feedback from students studying on the SPOC platform before class, observation of students' learning attitudes in flipped classrooms to display of students' results after class, and academic performance, summarize experience based on the analysis results, and optimize the teaching design plan. In classification algorithms, support vector machines (SVM) are widely used due to their advantages such as less overfitting and inconspicuous dimensionality of feature vectors. The traditional SVM algorithm is not suitable for processing large-scale data sets due to factors such as high time complexity and long training time. In order to solve these shortcomings, parallelizing the SVM algorithm to process large-scale data sets is an effective solution. On the basis of comparison, a SPOC-based flipped classroom teaching model was constructed, and empirical application was carried out in the Open University, in order to promote the sustainable development of open education.

## 1. Introduction

Big data technology is gradually playing an important role in various fields; it is also promoting the integration of information technology and courses to a new level [1]. As a novel information technology platform, MOOC has been widely used in the subject teaching of colleges and universities. Flipped classroom is a new teaching model that has been enriched and developed in the past educational theory and technology in recent years [2]. The quality education of graphic design in colleges and universities is an indispensable part of quality education. In recent years, it has received more

and more attention from colleges and universities [3]. Graphic design quality education in colleges and universities is to comprehensively improve the aesthetic ability of college students in the graphic design culture so that college students can feel the power of humanistic care and absorb the rich cultural, historical, and creative connotations contained in graphic design, thereby enriching college students' humanities feelings, to construct a diverse knowledge system of college students [4]. With the continuous development of informatization, the integration of information technology and graphic design quality education in colleges and universities has continued to deepen. Big data will usher in a new

era of comprehensive and deeper integration of information technology, education, and teaching. Educational informatization is also facing the transition from IT (information technology) to DT (data technology) [5]. From now on to the future, mobile Internet has deeply integrated people's lives, including education. From the perspective of the development of educational concepts, the characteristics of information technology better meet the needs of teaching students in accordance with their aptitude, personalized teaching, and paying attention to the development of each student [6].

Big data is a 4V (large volume, fast processing speed, many types, and great value) information technology recognized by the academia and the industry [7]. The generation of big data comes from the rapid development of information technology and the extensive application of the Internet, mobile Internet, and the Internet of Things. People and objects are called nodes that can be recorded and tracked in the networked society [8]. It is information technology that continues to integrate into people's lives. People and even objects will leave data traces in the information system, such as the use of social media, e-commerce, and online learning [9]. Fast processing speed means that the processing speed of big data is higher than in the past. Since many fields require near real-time data analysis, the storage and computing capabilities of large data are required to be higher than the previous data processing methods, but the storage and computing capabilities are large. While increasing the range, it also requires lower cost than the previous information system. Therefore, in the field of big data technology, in recent years, a scheme based on the Hadoop technology system has appeared, which basically meets the requirements of big data storage and computing [10]. Many types mean that the current data generated by various information systems is more similar to unstructured or semistructured data such as video, audio, pictures, and text. Therefore, big data technology needs to deal with unstructured data [11]. However, due to the huge amount of data and various types of data, data mining is needed to realize the value of the data [12].

This article is based on clarifying the related concepts of big data technology, MOOC, and flipped classroom. The relationship model between the three is proposed, that is, big data is the technical basis, MOOC is the application platform, and flipped classroom is the theoretical basis. This paper uses the support vector machine algorithm and big data technology to study the parallel SVM algorithm based on the Hadoop platform, analyzes the massive learning data in the MOOC platform, and dynamically predicts whether the learner can finally obtain the certificate during the learning process to help learners understand their own learning status, and the MOOC platform can also provide learners with customized learning planning and guidance based on this indicator to increase the rate of MOOC certificate acquisition and promote the development of the MOOC platform. At the same time, this article also proposes a model of integrated development among the three, in order to promote the practice of graphic design quality education in colleges and universities in the era of big data. The model divides the entire teaching design into two modules: SPOC platform teaching activity design and flipped classroom teaching

activity design. We apply this model to the actual teaching of open education to design detailed teaching activities. With reference to the above-mentioned understanding of SPOC, flipped classrooms, and MOOCs, this research applies the improvement and reshaping of MOOC resources to small-scale specific groups of people, which is supplemented by SPOC lecture videos and online interactive evaluation functions in traditional campus classrooms, through learners' personalized learning methods, accepts knowledge and meaning to construct new knowledge and integrate the process of online and offline interactive communication between teachers and students, and explores a flipped classroom with adaptive learning as a service and "individualized learning" as the concept teaching mode and teaching activity design.

## 2. Related Work

In terms of the construction of instructional design model, the current domestic and foreign researches on the construction of SPOC-based flipped classroom instructional design model are still lacking. More scholars are concerned about the connotation, characteristics, and characteristics of SPOC. In terms of teaching methods and other issues, there is still a lack of research on the promotion of open education learning under the flipped classroom environment of SPOC [13]. At the same time, the flipped classroom was rated as one of the top ten educational technology events by the famous online education media in the United States. In recent years, the flipped classroom teaching model has become a hot topic in the field of higher education and educational technology at home and abroad [14]. Professor Hew [15] believed that the new teaching model based on the flipped classroom model technology may be another major innovation after the class teaching model. In the United States, some colleges and universities and scientific research institutions have also carried out research on the domestic and foreign flipped classrooms and national flipped classroom teaching practices. However, there is still a lack of research on the construction of SPOC-based flipped classroom teaching design models. Classroom instruction design mode is a relatively new research focus. To solve these problems, Decheng and Jinxin [16] proposed SPOC, which was a small-scale private online course. They believed that MOOC is only a supplement to classroom teaching and cannot completely replace traditional teaching. When MOOC is developed to be able to contribute to teacher utilization, student output, student self-learning ability, and learning participation, MOOC will develop into SPOC. Stephany [17] found that from MOOC to SPOC, it provided resources, environment, and ideological support for students to transform from external learning to internalized learning. SPOC surpasses MOOC in terms of operation mechanism, teaching format, teaching process, and teaching structure. In terms of operating mechanism, a "small-scale (less than 500 students)" and "exclusive access condition" mechanism is proposed for online courses, which greatly reduces the complexity of learner management; it stimulates independent learning motivation, enhances learning interaction,

and improves the course completion rate and learning effect; in terms of teaching format, SPOC is committed to developing blended learning, which is beneficial to deeply integrate the advanced ideas of MOOC with traditional teaching.

Domestic scholars have different opinions on its use in teaching. King and others [18] believed that flipped classrooms were not only conducive to the humanization of education but also promote the harmonious reconstruction of teacher-student relations. However, there are also scholars who say that flipped classrooms are useful. There are few applications in the classrooms of primary and secondary schools, and they believe that this teaching method is not binding on young learners. However, in higher education and open education, it is more widely used for learners with learning autonomy. By summarizing the opinions of relevant scholars at home and abroad, it can be considered that flipped classroom is a learning and teaching model suitable for students who have an active learning attitude willing to take up their spare time and have basic self-study capabilities and self-study equipment that can watch online videos. Wu [19] believed that the evaluation of flipped classrooms should be completed by experts, teachers, peers, and learners themselves. Flipped classroom not only pays attention to the evaluation of learning results but also pays attention to the evaluation of the learning process through the establishment of student files of students, so as to achieve quantitative evaluation and qualitative evaluation, formative evaluation and summative evaluation, and individual evaluation and group evaluation. Zhang [20] pointed out the problems encountered in the practice of flipped classrooms: objective factors include the school's current teaching facilities and subject adaptability and subjective factors include teachers' beliefs in educational reforms and teachers' professional capabilities, and proposed solutions [21].

### 3. Construction of a Flipped Classroom Platform for Big Data Courses Based on Graphic Design MOOC

*3.1. Distribution of Graphic Design Modes.* The flipped classroom teaching model is divided into four stages. The first two stages are guided by teachers, and the latter two stages are actively carried out by students. Experiential engagement: it uses interactive forms such as games and simulations, experiments, community activities, and artistic activities. Conceptual connections: it provides video and audio courses, rich websites, and online discussions. Meaning making: it uses blogs, experiential videos, audiovisual reflections, and quizzes to present ideas. Demonstration and application: it includes creative and personal activities and displays. Figure 1 shows a schematic diagram of the graphic design pattern distribution.

Suppose that in system  $s$ , input  $x$  and output  $y$  have some unknown relationship, that is, there must be an uncertain probability  $F(x, y)$ , and an optimal function can be found in the function set according to the training samples.  $R$  represents the probability of output between system  $S$  and learner LM, that is, the probability of classification error.

Calculating the joint probability  $F(x, y)$ , you can find the minimized expected risk  $R$ . Since the joint probability  $F(x, y)$  is unknown,  $R$  cannot be directly calculated and minimized. However, using the theorem of large numbers, based on the known training sample set, an arithmetic average can be used.

$$A = \{(x(i), y(i)) \mid x(i), y(i) \in R\}, \quad i \in N. \quad (1)$$

When the dimension of the function set is relatively high, the empirical risk is small at this time, but the expected risk is not the smallest, which will make the generalization ability of the function set weaker, which is a phenomenon of overlearning. Principle of minimizing structural risk: we decompose the function set  $S$  into a sequence of function subsets, and then, the subset sequence has the following relationship:

$$f(x) - (a \times x(i) + b) = 0, \quad (2)$$

$$\begin{cases} \frac{\partial g(x, y)}{\partial x(i)} = 1, \\ \frac{\partial g(x, y)}{\partial y(i)} = 1. \end{cases} \quad (3)$$

For each subset, there is a certain confidence range, within which the function with the least empirical risk is searched. For different subsets, as long as the set with the smallest sum of the confidence range and the smallest empirical risk is selected, the subset of the function with the smallest expected risk can be obtained. Therefore, the principle of structural risk minimization refers to finding a balance between the generalization ability of the function set and the complexity contradiction.

$$|s(x, p)| = \sqrt[p]{x(1)^p + x(2)^p + \dots + x(n)^p}. \quad (4)$$

As the function subset continues to increase, its empirical risk continues to decrease, but the corresponding confidence range continues to grow. The structural risk minimization principle takes this situation into account so that the confidence range and empirical risk are selected in the function subset. In this way, the empirical risk is minimized and the best boundary is found, and the minimized expected risk can be obtained.

$$\begin{cases} t \times x(i) - r < 1, & i < t, \\ t \times x(i) - r > 1, & i > t. \end{cases} \quad (5)$$

Hadoop is a distributed system infrastructure, mainly centered on HDFS (Hadoop Distributed File System) distributed file system and MapReduce parallel computing framework. With the emergence of Hadoop, developers can use the advantages of clusters to write distributed applications that process massive amounts of data only by writing Map and Reduce functions, without needing to understand

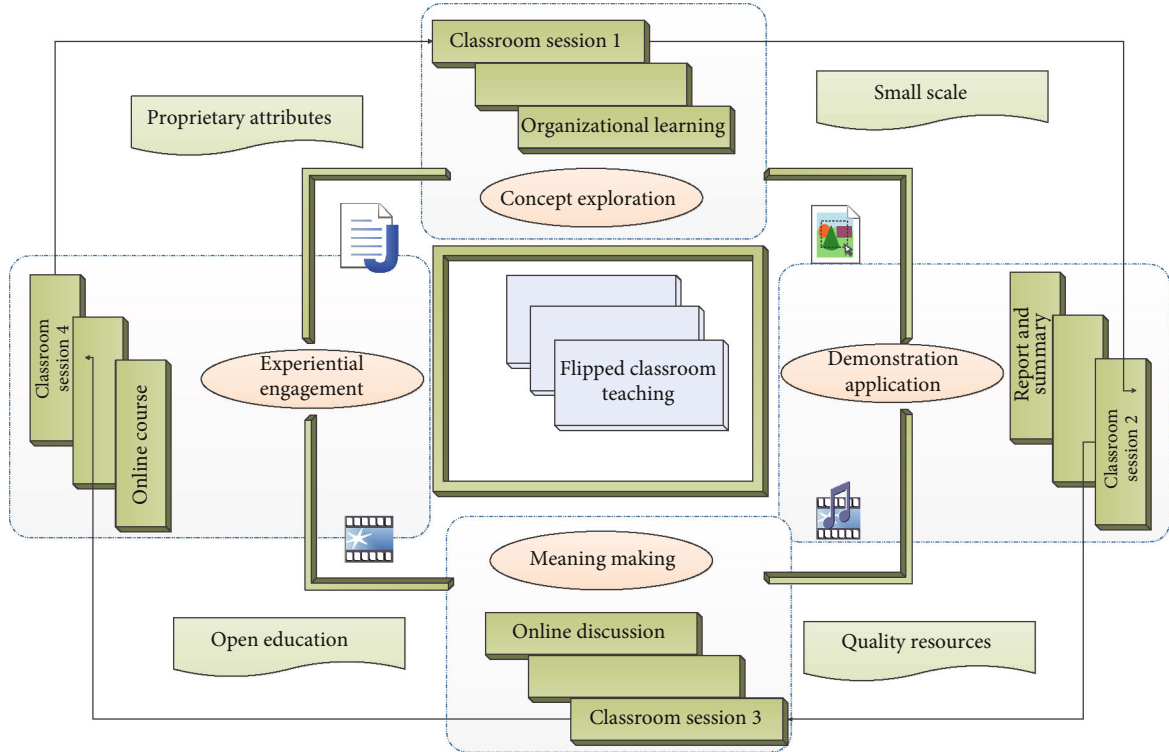


FIGURE 1: Schematic diagram of graphic design mode distribution.

the details of the underlying distributed network transmission and communication.

$$f(x(i)) \times (a \times x(i) + b) - 1 > 0, \quad \forall i = 1, 2, \dots, N. \quad (6)$$

Normally, the Secondary Node is in a dormant state. Once the Name Node goes down, the Secondary Node will be activated immediately and take over the Name Node to manage the entire cluster.

$$C = \frac{\sum_{i=1}^N \sum_{j=1}^N (a(i) - x(i)) \times (a(j) - x(j))}{\sum_{i=1}^N \sum_{j=1}^N (a(i) \times a(j))}. \quad (7)$$

Because the information of the Name Node changes dynamically with the entire cluster, the Name Node and the Secondary Node need to be synchronized to ensure that once a problem occurs on the Name Node, the Secondary Node can completely replace him. The actual operation of the client's data is to communicate with the Data Node through the TCP protocol. This design enables a Name Node to manage a large number of Data Nodes at the same time.

**3.2. Cross-Validation of Big Data Algorithms.** The curriculum system uses the machine learning method [22–24] of big data to understand the student's time schedule and arranges the graphic design courses inadvertently. It may be when eating in the cafeteria or between professional classes. Evaluation indicators are more diversified, and team learning has become an important part of the evaluation indicators. According to the student situation of the school,

the learning unit with strong relevance to the big data course is selected. The bottom layer of the framework is the Hadoop platform, including HDFS and MapReduce. The upper layer of the framework is the most commonly used classification, clustering, and association analysis algorithms for analyzing data. Classification analysis algorithms include support vector machines (SVM), decision trees, KNN and their corresponding parallel algorithms, and clustering analysis algorithms [25–27] including  $k$ -means; these algorithms include classic algorithms and the most popular new algorithms and have low time complexity and relatively good clustering results. Figure 2 shows the big data algorithm cross-validation flowchart. Each individual computing node still trains the subsample set separately. The difference is that the improved algorithm no longer merges the support vectors obtained from the corresponding training sample subset according to the two-by-two merge rule, but combines all nodes. The samples corresponding to the support vectors are extracted and combined, and then, the combined data set is divided into the number of computing nodes in the cluster and then handed over to all computing nodes in the cluster to train and learn individually. All subsequent layers follow the same rules. In this way, in the training and learning of each layer, the computing power of all nodes in the cluster is fully utilized, and no computing nodes are left in an empty state and because the data size after the block is compared with the previous SVM. All parts outside the first layer should be small, which also reduces the size of the corresponding data set, speeds up the training time of the SVM, and improves the convergence speed of the optimal solution.

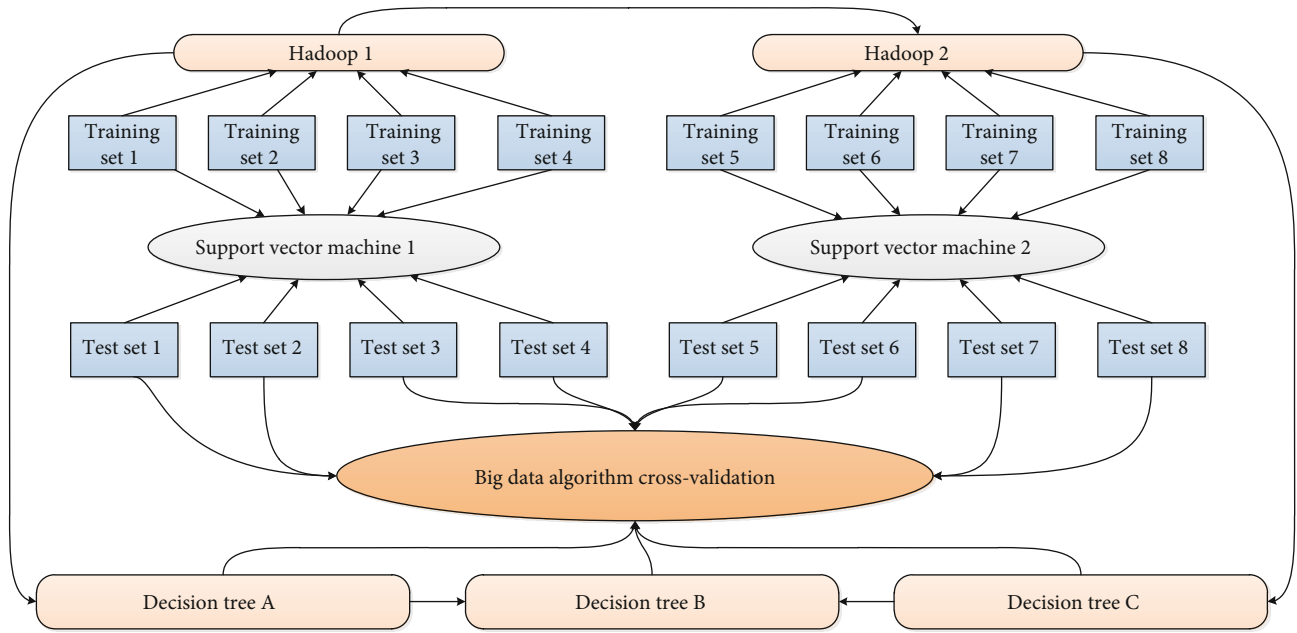


FIGURE 2: Big data algorithm cross-validation flowchart.

In order to improve the problems that exist when the number of positive and negative samples differs greatly, we adopt a weighted improvement method and use different penalty parameters for the positive and negative sample sets. This data set is mainly used to judge whether the learner can pass the MOOC exam and obtain the certificate of completion based on the 6 characteristics of the students' learning behavior on a domestic MOOC site. We can apply this predictive model to the entire learning process of MOOC learners and dynamically monitor the learning behavior of learners so that we can provide them with customized services during the learning process of MOOC learners. To remind and help them obtain the final study certificate, the main research content in this section is for the case where the SVM algorithm has a large difference in the size of the class sample in the two-class classification problem.

**3.3. Optimization of Flipped Classroom Evaluation.** In order to ensure the effectiveness of online courses, MOOC has made meticulous designs in all aspects of the course design. According to the design concept of MOOC, the scope of communication, the technology adopted, the applicable objects, and the design of teaching objectives, there are three main categories (cMOOC, xMOOC, and tMOOC). Among them, tMOOC adopts a task-based learning method, which is more suitable for graphic design quality education. The past graphic design quality courses such as graphic design appreciation and graphic design history are single and relatively boring. Using MOOC's task-based curriculum approach, graphic design appreciation can be transformed into learning team tasks, and the MOOC's more powerful platform communication capabilities can help each task group complete the learning tasks. In addition, during the learning task process, team and individual task reports and evaluations can be carried out in stages, and the primary

evaluation method can be transformed into a process evaluation method, which effectively guarantees the effectiveness of the curriculum. In the classroom, according to the pre-class online learning situation, teachers organize teaching activities in a targeted manner, answering questions and solving potential learning problems of learners and making innovative applications. Research shows that what students need most are not teachers' explanations of knowledge or concepts or colorful video learning, but when they encounter difficulties in practice, and flipped classrooms can solve this problem. Using the flipped classroom teaching model to focus valuable classroom time on active project-based learning, while the relevant knowledge and materials of the graphic design course itself require students to actively and autonomously learn after class, teachers use various informatization to learn independently and students can also fully interact on the information platform.

Figure 3 shows the histogram of the flipped classroom content design unit. The content design of video teaching includes two units: Hadoop platform construction (application part) and common algorithms for big data processing and analysis (theoretical part). The application part includes 8 videos, and the theoretical part includes 10 videos. The length of the video is between 10 and 15 minutes. When processing data, the data will be divided into multiple blocks according to the data to be processed and the function written by the developer, and then, a map task will be started for each block. After being scheduled by the MapReduce runtime environment, it runs on several nodes in the cluster. When all map operations in the cluster are executed, many intermediate key-value pairs will be generated. The intermediate key-value pairs will be redivided into multiple groups by a partitioner component. The MapReduce runtime environment will schedule a reduce task to calculate each one. After all reduce functions are executed, all calculation results

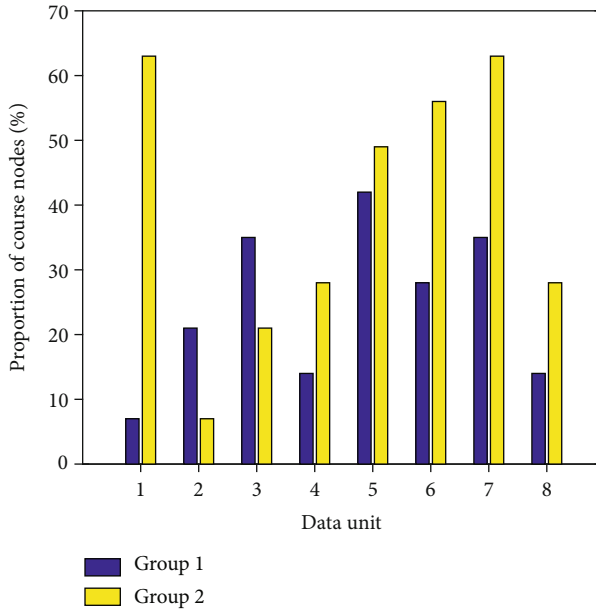


FIGURE 3: Histogram of flipped classroom content design unit.

will be output to the specified location on HDFS, and the client can read the calculation results from it. Usually, in a MapReduce cluster, the node that receives the tasks submitted by the client is called the master node, and there is a JobTracker process on the master node that is responsible for receiving specific tasks. The node responsible for executing map or reduce tasks is called the slave node, and there is a TaskTracker process running on the slave node that is responsible for executing specific tasks. The MapReduce program usually consists of mapper code, reducer code, and related configuration parameters.

#### 4. Application and Analysis of Flipped Classroom Platform for Big Data Courses Based on Graphic Design MOOC

**4.1. Graphic Design MOOC Index Processing.** The experimental CPU model is Intel i3-3110M, the memory is Kingston 1333 Hz, the capacity is 2 G, and Matlab2012. The SVM algorithm used is the Libsvm-3.20 version toolkit. This toolkit has been since version 2.8. The SVM training algorithm uses an optimized SMO algorithm. Teaching evaluation is an important way to test the achievement of teaching results and course objectives, and it is also an important means to evaluate the teaching quality of big data courses. This data set is mainly based on 6 characteristics of students' learning behavior on a domestic MOOC site to determine whether a learner can finally pass the MOOC exam and get a certificate of completion. Figure 4 shows the level distribution of the flipped classroom evaluation pyramid. The six learning behavior features extracted in this article are as follows: total number of times to watch videos, total number of quizzes submitted, recording density, number of posts in forums, number of times to see posts on forums, and how many days to register before the start of

class course. The main extractions are data structure, machine learning [28–30], and advanced English writing three MOOC courses. The specific description of the data set is as follows: the number of sample points of the data structure course  $m = 85302$ , the number of sample points of the machine learning course  $m = 75791$ , and the number of sample points of the advanced English writing course [31, 32]  $m = 61218$ ; the feature dimension of the sample points  $n = 6$ , the learners who have obtained the certificate are in the positive category, those who have not obtained the certificate are in the negative category, and the number of categories  $s = 2$ .

In this paper, different learning weights are assigned to the learning trajectory of students on the school network platform, combined with the characteristics of big data courses, and the teacher and students jointly score the learning situation of each student. For example, the construction of the Hadoop platform is scored by the teacher. The learning situation after class is scored by the students, and the total score of each student is calculated according to the weight as the usual score of the course and recorded in the student's score. A set of formative evaluation-oriented assessment methods are designed: class attendance rate 15%, online test 5%, group presentation 15%, analysis report 15%, and final exam 50%. Figure 5 shows the pie chart of the flipped classroom evaluation method.

The teacher team uploads the microvideos and other resources made by SVM to the network platform for students to download and watch. After the students complete their self-study, they complete the online self-test assignments. The size of the group is controlled within 5 people, and each group of students selects part of the content for design, and the time is 20 minutes. The SVM algorithm is more complicated, and the problem can be divided into several subproblems, such as discussing the main idea and implementation process of the SVM algorithm. As for the advantages and disadvantages, the students carry out division of labor and cooperation, and the teachers group to guide and communicate. According to the students' understanding of the main idea of the SVM algorithm, the SVM algorithm is realized through different programming languages. After the algorithm is completed, a representative is selected to explain the algorithm completion process and answer the teacher and other students' questions. In this way, the unity and cooperation between students and the communication between teachers and students can be enhanced. The integration of blended learning mainly includes preevaluation of the effectiveness of SPOC teaching resources, trial and evaluation of the comprehensiveness of the SPOC platform and teaching service system, recognition of teaching innovation, school incentive policies, and student evaluation system. After class, students will complete the SVM algorithm in the process of implementing the SVM algorithm on the school network platform and other members of the group. The weight is scored, and the teacher scores the learning situation of each student.

**4.2. Flip Classroom Platform Simulation.** Before implementing flipped classroom teaching, first we select teaching



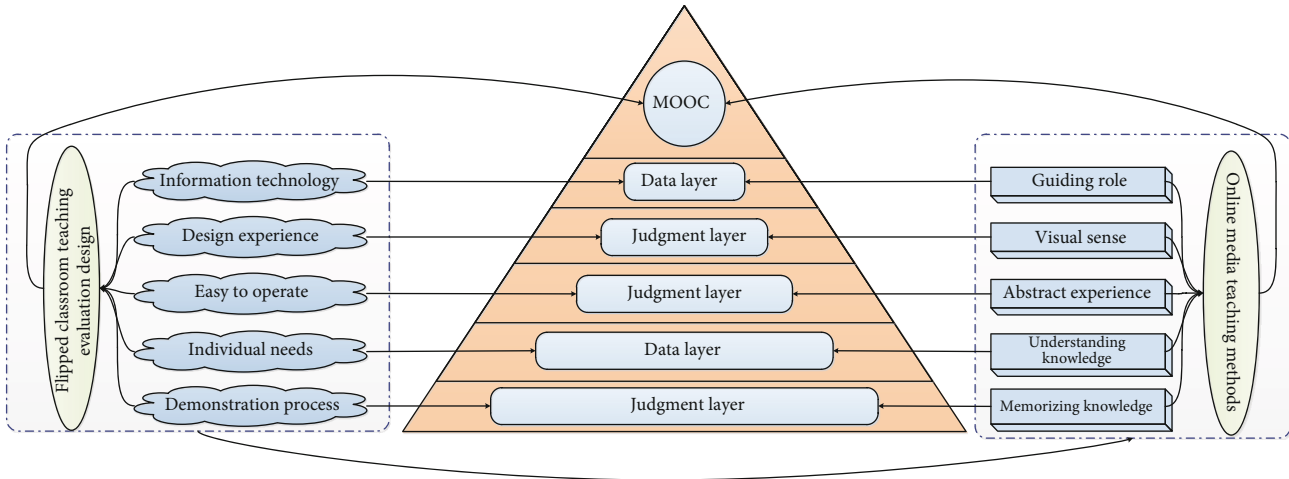


FIGURE 4: Flipped classroom evaluation pyramid level distribution.

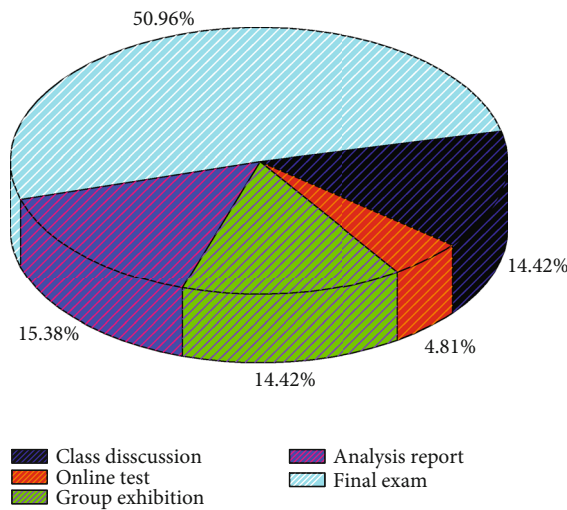


FIGURE 5: Fan-shaped chart of flipped classroom evaluation assessment method.

content, design questions and resources that can inspire students to explore, formulate a teaching list, and provide students with independent learning. In addition, two classes were selected for comparative teaching, hoping to see different teaching effects; representative content was selected for video teaching and related resources and questions based on textbooks and network resources. In addition, two parallel classes of 16 levels (45 students in each class) were selected to carry out comparative experiments. Among them, it adopts flipped classroom teaching, and another adopts traditional teaching mode. Figure 6 shows the division of design elements for flipped classroom teaching content.

Before the questionnaire was officially issued, the students were tested, some items were removed, and 19 formal items were finally determined, including changing concepts and the popularization and promotion of MOOC, hardware and teaching environment, establishing a high-quality MOOC teaching platform. The questionnaire consists of 5 aspects and is divided into 5 items. Each item uses a 5-

point scoring method, which is divided into “completely agree,” “relatively agree,” “neutral attitude,” and “disagree.” There are five levels of “complete disagreement,” which are divided into 1, 2, 3, 4, and 5 in turn. Figure 7 shows the evaluation results of the MOOC-based flipped classroom teaching process. According to factor analysis, the questionnaire is divided into 7 factors: establishment of a high-quality MOOC teaching platform, reform of education and teaching management system, teachers’ work before flipped classroom, classroom teaching of flipped classroom, teacher quality, students’ independent learning ability, and student information technology literacy.

This survey focuses on the students and teachers of adult higher education in the main and branch campuses of the college. In the end, 248 student questionnaires were recovered, 211 valid questionnaires were returned, and the effective response rate was 85%. 20 teacher questionnaires and 18 valid questionnaires were recovered, and the effective recovery rate was 90%. It can be seen that 46.4% of the students hope that the teacher will assign the learning tasks of the MOOC course before the MOOC-based adult higher education flipped classroom class; 59.2% of the students hope that the teacher will assign the learning tasks of the learning materials. The teacher assigned the most recommended books and materials to the students, reaching 68.7%; 52.1% of the students hope the teacher assigned to complete the basic homework. It can be seen that the students hope that the teacher can arrange and study MOOC courses, textbooks, recommended books, and materials and complete the basic tasks before class.

**4.3. Example Application and Analysis.** In the classroom session, the first is the configuration of the project environment, preparing multiple hosts for students to run HDFS. Each host needs to be installed: Linux operating system, Hadoop-2. 8.0, jdk-8u131-linux-64, and Hadoop Eclipse plug-in. The specific process includes connecting to the network, installing JDK, creating HDFS account, configuring password-free SSH access mechanism, installing Hadoop and configuring environment variables, modifying Hadoop

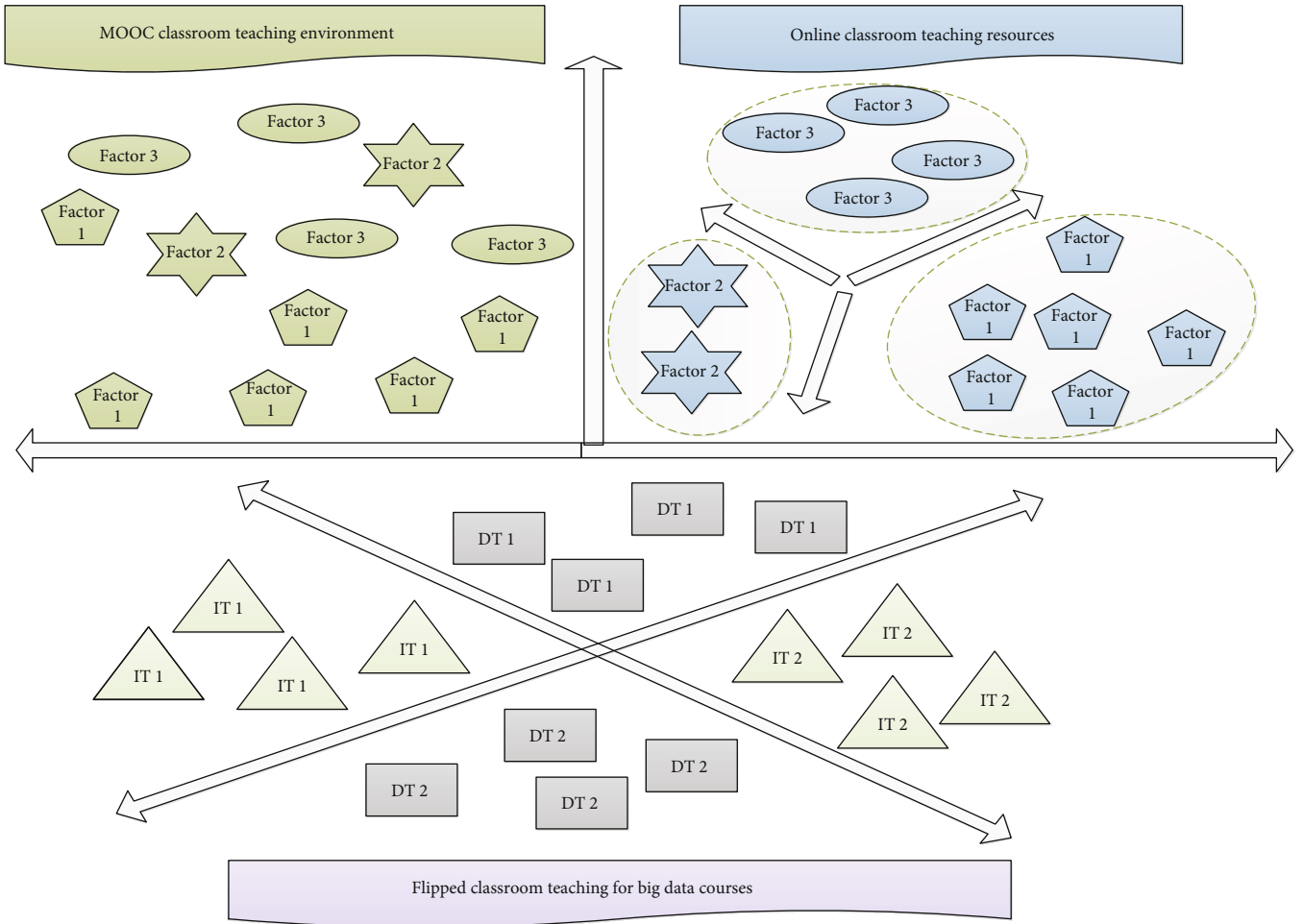


FIGURE 6: The division of design elements of flipped classroom teaching content.

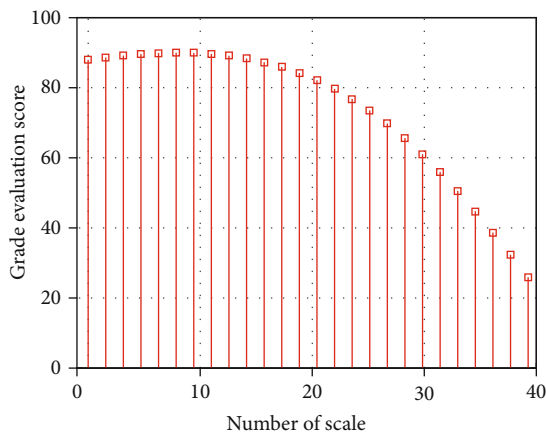


FIGURE 7: Evaluation results of the teaching process of the flipped classroom based on MOOC.

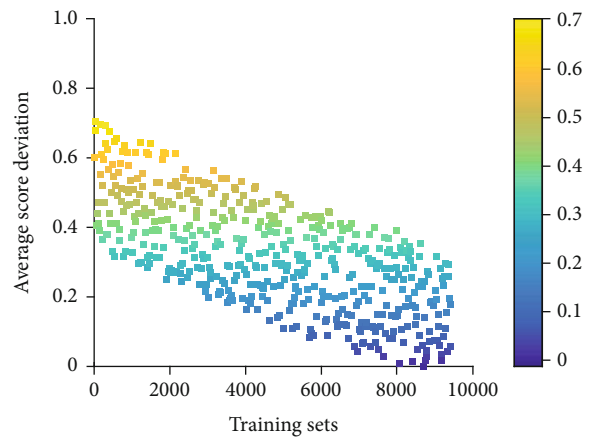


FIGURE 8: Two-dimensional scatter plot of weight score deviation under different training sets of the model.

configuration files on Name Node, and copying Hadoop system to all Data Node. The software and the construction process of the cluster environment are placed on the school network platform as teaching resources, and the students build the cluster environment in groups. For the problems encountered by the students during the environment con-

struction process, the group communicates with each other, and teachers can guide as needed. There are four ways to evaluate SPOC-based flipped classroom teaching effects: in-class test evaluation, homework evaluation, peer evaluation, and assessment evaluation. For learners, they are more inclined to use in-class test evaluation and comprehensive

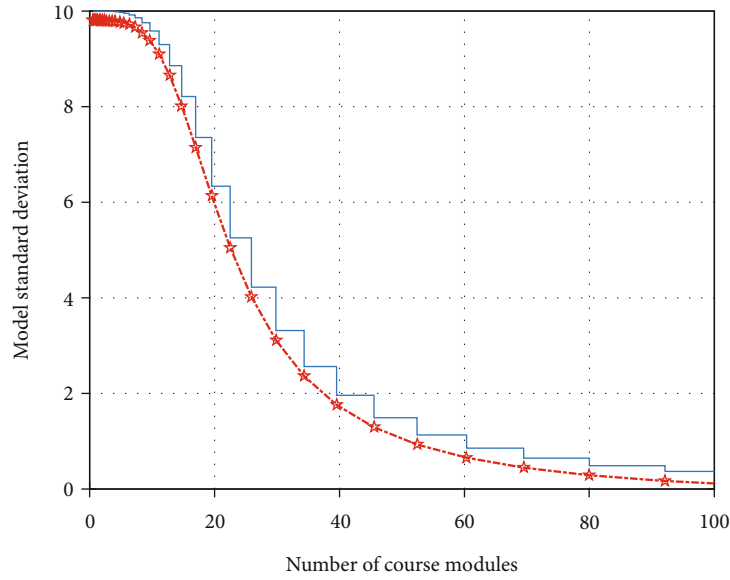


FIGURE 9: Model evaluation statistical standard deviation ladder diagram.

evaluation. The in-class test evaluation can more intuitively detect your own learning effect, and the comprehensive evaluation can more comprehensively understand your own learning effect. After class, teachers and students on the school network platform combined with other members of the team jointly score the performance of the HDFS construction process according to different weights. Figure 8 shows the weight scoring deviation of the model under different training sets' two-dimensional scatter plot.

The SPOC platform discussion area is divided into 3 submodules: teacher Q&A area, classroom communication area, and comprehensive discussion area. From it, we can see that the number of participants in the teacher's Q&A area is small, the number of people in a single day is basically below 100, and the daily number of people fluctuates widely. It shows the relevant situation of the top seven topics in the classroom exchange area. We can see that 3 of these 7 topics are attended by teachers, the number of views reached 4012, the number of replies was 4796, and 482 votes were voted. The seventh-ranked topic was Discussion 1, and the number of views reached 9474. The number of replies was 7937, and 193 votes were cast. The change reflects the enthusiastic discussion in the classroom exchange area, and the topic content is more inclined to the Internet.

Figure 9 shows a ladder diagram of the statistical standard deviation of the model evaluation. We can see that in terms of online learning, the highest score of students is 100 points, the lowest score is 4 points, the average score is 84.15, and the standard deviation is 5.661, indicating that the online learning scores of the studied student groups are quite different. Online learning is not a necessary learning link, it does not have much appeal to students, and it is optional and only part of the course videos is watched. Regarding online discussion, the highest score of students is 100, the lowest score is 75, the average score is 90.15, and the standard deviation is 5.161, indicating that the differences in the scores of online discussions among the stud-

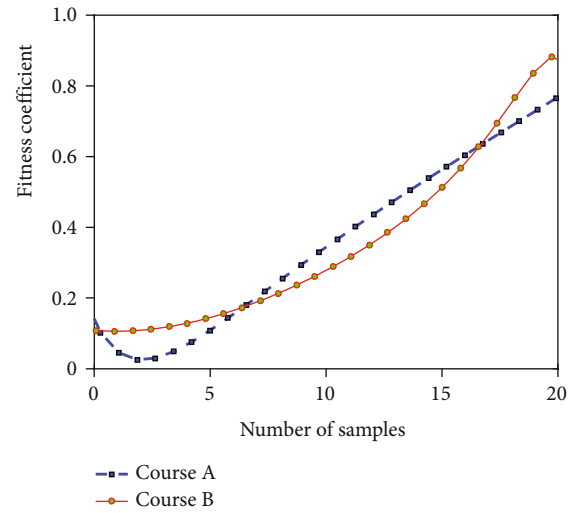


FIGURE 10: Curves of fitness coefficients of different courses.

ied student groups are small, which means that students generally like online discussions and are willing to invest more time to the online discussion. Otherwise, we can see that 76.03% of the students agree with "you like the flipped classroom teaching model used in this course," but only 8.57% disagree, which shows that this teaching model is very popular among students. From that, we can see that 68.79% of the students agreed with "this model can improve learning efficiency," while only 2.49% disagreed.

Figure 10 shows the fitness coefficient curves of different courses. We can see that as the number of samples increases, the value of the best fitness keeps increasing until the number of iterations exceeds 20 and the algorithm ends. As the number of samples increases, the average fitness shows a trend of substantial increase. The classification result of the training set is 85.1823%. According to the best parameters selected by genetic algorithm  $C = 1.2871$ ,  $g = 6.3459$ , tested

in the test set, the classification accuracy rate obtained is 84.7671%. The average fitness also decreases slightly at the beginning of the algorithm and then gradually stabilizes. The classification result on the training set is 85.1531%. According to the best parameters selected by the particle swarm algorithm  $C=9.6356$ ,  $g=0.5471$ , after testing on the test set, the classification accuracy rate is 86.0152%. Through the comparison of experiments, it can be seen that the classification accuracy of the particle swarm algorithm is slightly higher than that of the genetic algorithm on the training set and the test set. It can be seen that the number of people has shown an obvious trend of change with the development of the course. At the beginning, students are more motivated to learn, but gradually, the number of people watching videos is getting less and less, reflecting that students' learning initiative is getting weaker and weaker. The convergence speed of the algorithm is faster than that of the other algorithm.

## 5. Conclusion

This research combines MOOC and SPOC to promote the transformation of college classroom teaching mode in the form of flipped classrooms. This transformation enables MOOC to be implemented in college physical classrooms. At the same time, MOOC and SPOC provide high-quality preclass learning for the implementation of flipped classrooms. In this paper, combining the Hadoop platform that has been widely used in recent years, using the advantages of Hadoop's ability to effectively store massive data and natural support for parallel computing, a parallel SVM algorithm based on Hadoop is designed, which greatly improves the training speed of SVM and uses two heuristic parameter optimization strategies (particle swarm algorithm and genetic algorithm) to optimize SVM parameters so that the classification accuracy rate has been improved to a certain extent. Experiments have verified the effectiveness of the algorithm, and in view of the uneven characteristics of the sample data of the MOOC platform, a parameter weighting solution is proposed; the availability of the processing method is also confirmed in the experiment. Based on the results of SPSS software data analysis, this paper proposes that the implementation of MOOC-based higher education flipped classrooms requires comprehensive consideration of concept changes and in-depth analysis of the feasibility of implementing MOOC-based higher education flipped classrooms. The generalization ability of the classifier and the improvement of classification accuracy are obtained. This significantly improves the training speed of the SVM algorithm when processing massive amounts of data and at the same time ensures a better classification accuracy, and in view of its shortcomings, two suggestions for improvement are proposed and applied to processing large-scale data sets, which greatly improves the training speed of SVM.

## Data Availability

The data used to support the findings of this study are included within the article.

## Conflicts of Interest

The author does not have any possible conflicts of interest.

## References

- [1] W. Zhang and E. Zhang, "RETRACTED: On China's English teaching reform: from micro-lesson and MOOCs to flipped class," *International Journal of Emerging Technologies in Learning*, vol. 13, no. 1, pp. 220–229, 2018.
- [2] P. J. Muñoz-Merino, J. A. Ruipérez-Valiente, C. D. Kloos et al., "Flipping the classroom to improve learning with MOOCs technology," *Computer Applications in Engineering Education*, vol. 25, no. 1, pp. 15–25, 2017.
- [3] Y. C. Chen, K. K. Fan, and K. T. Fang, "Effect of flipped teaching on cognitive load level with mobile devices: the case of a graphic design course," *Sustainability*, vol. 13, no. 13, 2021.
- [4] G. Sammour, A. Al-Zoubi, and J. Schreurs, "Opportunities of MOOCs and flipping micro-learning models in international joint academic degree programs," *International Journal of Technology Enhanced Learning*, vol. 12, no. 4, pp. 411–425, 2020.
- [5] J. C. H. Huang, A. Y. Q. Huang, and O. H. T. Lu, "Learning dashboard: visualization of learning behavior in MOOCs," *Technology-Enhanced Collaborative Learning*, vol. 1, pp. 21–25, 2018.
- [6] B. Prevalla and H. Uzunboylu, "Flipped learning in engineering education," *TEM Journal*, vol. 8, no. 2, 2019.
- [7] B. Y. Chen, D. E. Kern, R. M. Kearns, P. A. Thomas, M. T. Hughes, and S. Tackett, "From modules to MOOCs: application of the six-step approach to online curriculum development for medical education," *Academic Medicine*, vol. 94, no. 5, pp. 678–685, 2019.
- [8] R. R. de Castañeda, A. Garrison, P. Haerberli et al., "First 'Global Flipped Classroom in One Health': from MOOCs to research on real world challenges," *One Health*, vol. 5, pp. 37–39, 2018.
- [9] S. J. H. Yang and C. S. J. Huang, "Taiwan digital learning initiative and big data analytics in education cloud," *Advanced Applied Informatics*, vol. 6, pp. 366–370, 2019.
- [10] M. Chen, B. Li, and Y. Hu, "A teaching ecosystem of engineering graphics based on the deeper learning cycle model and asynchronous SPOC," *Journal for Geometry and Graphics*, vol. 23, no. 2, pp. 259–267, 2019.
- [11] A. P. Montgomery, D. V. Hayward, W. Dunn, M. Carbonaro, and C. G. Amrhein, "Blending for student engagement: lessons learned for MOOCs and beyond," *Australasian Journal of Educational Technology*, vol. 31, no. 6, pp. 11–14, 2019.
- [12] J. Qin and Z. Jia, "Learner-generated behaviours in a flipped learning: a focus on computer culture foundation course," *International Journal of Continuing Engineering Education and Life Long Learning*, vol. 30, no. 1, pp. 1–14, 2020.
- [13] M. V. Voronina, O. N. Moroz, A. E. Sudarikov, M. B. Rakhimzhanova, and E. K. Muratbakeev, "Systematic review and results of the experiment of a flipped learning model for the courses of descriptive geometry, engineering and computer graphics, computer geometry," *Eurasia journal of mathematics, science and technology education*, vol. 13, no. 8, pp. 4831–4845, 2017.
- [14] Y. Liu, X. Guo, and G. Li, "Research on deepening Python case teaching by flipped classroom method," *Distance Education and Learning*, vol. 2, pp. 107–111, 2020.

- [15] K. F. Hew, "Promoting engagement in online courses: what strategies can we learn from three highly rated MOOCs," *British Journal of Educational Technology*, vol. 47, no. 2, pp. 320–341, 2019.
- [16] Z. Decheng and C. Jinxin, "Curriculum analysis based on Chinese university MOOC platform," *Education journal*, vol. 7, no. 3, pp. 68–74, 2018.
- [17] F. Stephany, "One size does not fit all: constructing complementary digital reskilling strategies using online labour market data," *Big Data & Society*, vol. 8, no. 1, article 205395172110031, 2021.
- [18] C. King, J. A. Kelder, and K. Doherty, "Designing for quality: the understanding dementia MOOC," *Lead Issues elearning*, vol. 2, 2015.
- [19] L. Wu, "On blended learning mode of comprehensive English based on MOOC," *Big Data Analytics for Cyber-Physical-Systems*, vol. 1, pp. 915–919, 2019.
- [20] K. Zhang, "Mining data from Weibo to WeChat: a comparative case study of MOOC communities on social media in China," *International Journal on E-Learning*, vol. 14, no. 3, pp. 305–329, 2019.
- [21] C. STÖHR, "STEM MOOCs in practice—experiences from Chalmers X first MOOC on graphene science and technology," *Research Track*, vol. 6, 2019.
- [22] W. Chu, *Studies on the Effects of Wiring Density on Chip Package Interaction and Design Optimization with Machine Learning (Doctoral Dissertation)*, The University of Texas Libraries, 2021.
- [23] W. Sun, P. Zhang, Z. Wang, and D. Li, "Prediction of cardiovascular diseases based on machine learning," *ASP Transactions on Internet of Things*, vol. 1, no. 1, pp. 30–35, 2021.
- [24] C. Ying, Y. Shuyu, L. Jing, D. Lin, and Q. Qi, "Errors of machine translation of terminology in the patent text from English into Chinese," *ASP Transactions on Computers*, vol. 1, no. 1, pp. 12–17, 2021.
- [25] J. Chen, C. Du, Y. Zhang, P. Han, and W. Wei, "A clustering-based coverage path planning method for autonomous heterogeneous UAVs," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–11, 2021.
- [26] Y. Jiang, X. Gu, D. Wu, W. Hang, J. Xue, and S. Qiu, "A novel negative-transfer-resistant fuzzy clustering model with a shared cross-domain transfer latent space and its application to brain CT image segmentation," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 18, no. 1, pp. 40–52, 2021.
- [27] W. Chu, P. S. Ho, and W. Li, "An adaptive machine learning method based on finite element analysis for ultra low-k chip package design," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, p. 1, 2021.
- [28] M. Gao, W. Cai, and R. Liu, "AGTH-Net: attention-based graph convolution-guided third-order hourglass network for sports video classification," *Journal of Healthcare Engineering*, vol. 2021, 10 pages, 2021.
- [29] S. Qi, X. Ning, G. Yang et al., "Review of multi-view 3D object recognition methods based on deep learning," *Displays*, vol. 69, p. 102053, 2021.
- [30] J. Zhang, W. Wang, C. Lu, J. Wang, and A. K. Sangaiah, "Light-weight deep network for traffic sign classification," *Annals of Telecommunications*, vol. 75, no. 7-8, pp. 369–379, 2020.
- [31] L. Liang, Q. Yin, and C. Shi, "Exploring proper names online and its application in English teaching in university," *ASP Transactions on Computers*, vol. 1, no. 1, pp. 24–29, 2021.
- [32] L. Calimeris, "Effects of flipping the principles of microeconomics class: does scheduling matter?," *International Review of Economics Education*, vol. 29, pp. 29–43, 2018.

## Research Article

# Feasibility of Using Improved Convolutional Neural Network to Classify BI-RADS 4 Breast Lesions: Compare Deep Learning Features of the Lesion Itself and the Minimum Bounding Cube of Lesion

Meihong Sheng <sup>1,2</sup>, Weixia Tang,<sup>2</sup> Jiahuan Tang,<sup>2</sup> Ming Zhang,<sup>2</sup> Shenchu Gong,<sup>2</sup> and Wei Xing <sup>1</sup>

<sup>1</sup>Department of Imaging, The Third Affiliated Hospital of Soochow University, Changzhou 213003, Jiangsu, China

<sup>2</sup>Department of Imaging, Affiliated Hospital 2 of Nantong University, Nantong First People's Hospital, Nantong 226001, Jiangsu, China

Correspondence should be addressed to Wei Xing; [suzhxingwei@suda.edu.cn](mailto:suzhxingwei@suda.edu.cn)

Received 8 July 2021; Revised 3 August 2021; Accepted 6 August 2021; Published 8 September 2021

Academic Editor: Khin wee Lai

Copyright © 2021 Meihong Sheng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

To determine the feasibility of using a deep learning (DL) approach to identify benign and malignant BI-RADS 4 lesions with preoperative breast DCE-MRI images and compare two 3D segmentation methods. The patients admitted from January 2014 to October 2020 were retrospectively analyzed. Breast MRI examination was performed before surgical resection or biopsy, and the masses were classified as BI-RADS 4. The first postcontrast images of DCE-MRI T1WI sequence were selected. There were two 3D segmentation methods for the lesions, one was manual segmentation along the edge of the lesion slice by slice, and the other was the minimum bounding cube of the lesion. Then, DL feature extraction was carried out; the pixel values of the image data are normalized to 0-1 range. The model was established based on the blueprint of the classic residual network ResNet50, retaining its residual module and improved 2D convolution module to 3D. At the same time, an attention mechanism was added to transform the attention mechanism module, which only fit the 2D image convolution module, into a 3D-Convolutional Block Attention Module (CBAM) to adapt to 3D-MRI. After the last CBAM, the algorithm stretches the output high-dimensional features into a one-dimensional vector and connects 2 fully connected slices, before finally setting two output results (P1, P2), which, respectively, represent the probability of benign and malignant lesions. Accuracy, sensitivity, specificity, negative predictive value, positive predictive value, the recall rate and area under the ROC curve (AUC) were used as evaluation indicators. A total of 203 patients were enrolled, with 207 mass lesions including 101 benign lesions and 106 malignant lesions. The data set was divided into the training set ( $n = 145$ ), the validation set ( $n = 22$ ), and the test set ( $n = 40$ ) at the ratio of 7 : 1 : 2; fivefold cross-validation was performed. The mean AUC based on the minimum bounding cube of lesion and the 3D-ROI of lesion itself were 0.827 and 0.799, the accuracy was 78.54% and 74.63%, the sensitivity was 78.85% and 83.65%, the specificity was 78.22% and 65.35%, the NPV was 78.85% and 71.31%, the PPV was 78.22% and 79.52%, the recall rate was 78.85% and 83.65%, respectively. There was no statistical difference in AUC based on the lesion itself model and the minimum bounding cube model ( $Z = 0.771$ ,  $p = 0.4408$ ). The minimum bounding cube based on the edge of the lesion showed higher accuracy, specificity, and lower recall rate in identifying benign and malignant lesions. Based on the lesion 3D-ROI segmentation using a minimum bounding cube can more effectively reflect the information of the lesion itself and the surrounding tissues. Its DL model performs better than the lesion itself. Using the DL approach with a 3D attention mechanism based on ResNet50 to identify benign and malignant BI-RADS 4 lesions was feasible.

## 1. Introduction

Breast cancer is a serious threat to women's health and has become the world's most common cancer [1]. Early detection, early diagnosis, and early treatment can improve both survival and prognosis of breast cancer patients [2–4]. Greenwood et al. [5] have reported that breast MRI plays an important role in screening and assessing the extent of ductal carcinoma in situ (DCIS) and predicting the potential invasiveness. The degree of early enhancement reflects the vascular richness and blood perfusion of the lesion and can reflect the characteristics of the lesion. According to the guideline of the American College of Radiology (ACR), the possibility range of the BI-RADS 4 of malignancy is 2%-95% as defined by the breast imaging report and data system (BI-RADS) [6]. Lesions with BI-RADS 4 classification are difficult to define clearly. The signs of the lesions are overlapping and intricate. These lesions, benign or malignant, are all classified as BI-RADS 4, along with recommended invasive procedures such as needle biopsy to obtain pathological evidence [7–9]. Therefore, comprehensive understanding and improved evaluation methods of benign and malignant breast lesions are urgently needed to reduce invasive operations and the burden on patients.

In recent years, with the rapid development of artificial intelligence-assisted diagnosis systems, deep learning has emerged as a subfield of machine learning [10–13]. Its application in medical imaging has attracted much attention, along with its wide use in image recognition, segmentation, and analysis [14]. Several studies [15, 16] have attempted to increase the number of layers of CNNs from the original 5 layers of the AlexNet network [17] to the 19 layers of the VGG network. Theoretically, a deeper network leads to better effect, but the increase in network depth will also bring additional problems that in turn cause reduced performance. The main reason for the performance reduction was gradient dispersion (vanishing gradients in backpropagation lead to weakened error signal) and gradient explosion (accumulation of large error gradients results in infinity in loss function) that were caused by the increase in the number of network layers. The residual module was proposed by Khalili and Wong [15], which could effectively solve the aforementioned problems above and has become the standard configuration of CNNs.

The CNNs learned a large number of features. Some features were not important for the final result, while some others played a key role in predicting results thus deserve more attention. Based on this theory, Woo et al. [18] proposed the Convolutional Block Attention Module (CBAM). The so-called greater attention was to give higher weight to those key features. In this study, the efficiency of feature extraction and classification of BI-RADS 4 breast lesions with two segmentation methods was compared by the DL model with a 3D attention mechanism, so as to verify the feasibility of using an improved convolutional neural network.

## 2. Materials and Methods

*2.1. Study Cohort and Imaging Protocol.* The patients who underwent breast MRI examinations at Nantong First Peo-

ple's Hospital were retrospectively collected from January 2014 to October 2020. A total of 296 patients with breast lesions were enrolled in the study. Inclusion criteria: (1) the diameter of the lesion was greater than 1 cm, or lesions were visible to naked eyes at least two consecutive slices; (2) the image quality was high without obvious artifacts or distortion; (3) the lesions were all mass and showed irregular margins, or inhomogeneous enhancement, or ring enhancement in MRI and classified as BI-RADS 4 by the radiologist. Exclusion criteria: (1) the breast mass showed no enhancement; (2) radiotherapy/chemotherapy or invasive operations such as biopsy before breast MRI; (3) the characteristics of the lesion and the pathological diagnosis were not clear.

All MRIs in this study were acquired using a Siemens 3.0 T magnetic resonance scanner (Verio; Siemens, Erlangen, Germany) with 16-channel phased array breast-specific coil. The patients were placed in the prone position with head-first entry; the breasts naturally hanged in the breast coil, and the nipple remained at the center of the coil. The scan sequence parameters were as follows: DCE T1-weighted axial fat suppression 3D spoiler gradient echo: TR 4.67 ms, TE 1.66 ms, flip angle 10°, FOV 340 mm × 340 mm, slice thickness 1.2 mm, scanning of 6 phases without interval, scan time 6 min 25 s, high-pressure syringe injection of 15–20 mL contrast agent Gd-DTPA based on body weight (0.2 ml/kg) at a flow rate of 2 mL/s, and then injection of the same amount of normal saline to flush the tube. After the 25 s injection, scanning was triggered, and each phase was collected for 1 min. The first phase was nonenhancement, and phases 2–6 were enhanced. Our study focused on phase 2 images which was named DCE-MRI T1WI first postcontrast sequence.

*2.2. 3D-ROI Lesion Segmentation.* All DCE-MRI T1WI first postcontrast images of breast mass that meet the inclusion criteria were imported into the image processing software ITK-SNAP 3.8.0 in DICOM format, and the lesions were manually segmented by an attending physician with 8 years of experience in breast MRI diagnosis and reviewed by a chief physician with more than 10 years of experience in breast MRI diagnosis: (1) based on the ROI of the lesion itself (Figures 1 and 2), the 3D-ROI segmentation method was used to manually delineate the boundary of the lesion slice by slice along the edge of the lesion, containing cystic degeneration, necrosis, and calcification within the lesion; (2) based on the minimum bounding cube, the maximum diameter of the lesion was then projected onto 3 coordinate axes of the image to determine its coverage range of  $x$ ,  $y$ , and  $z$  axes, and the bounding box of the lesion was finally obtained (Figures 3 and 4).

*2.3. Lesion Feature Extraction.* There are two methods of feature extraction. One is to take the minimum bounding cube of the lesion (including the lesion and part of the peritumoral area), and the other is to take only the lesion itself and set the value of the image pixels of part of the nonlesion area to 0. The minimum bounding cube is the smallest circumscribed cube containing the lesion. In addition, before

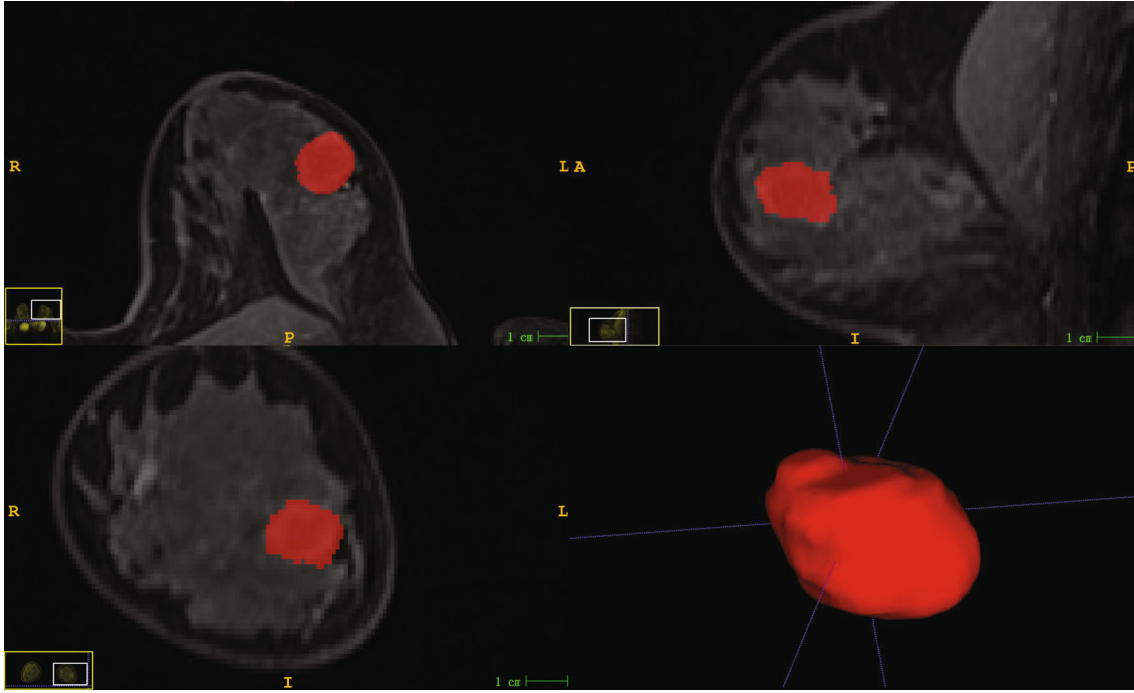


FIGURE 1: 3D-ROI segmentation method based on the lesion itself. Shows a case of fibroadenoma of the left breast.

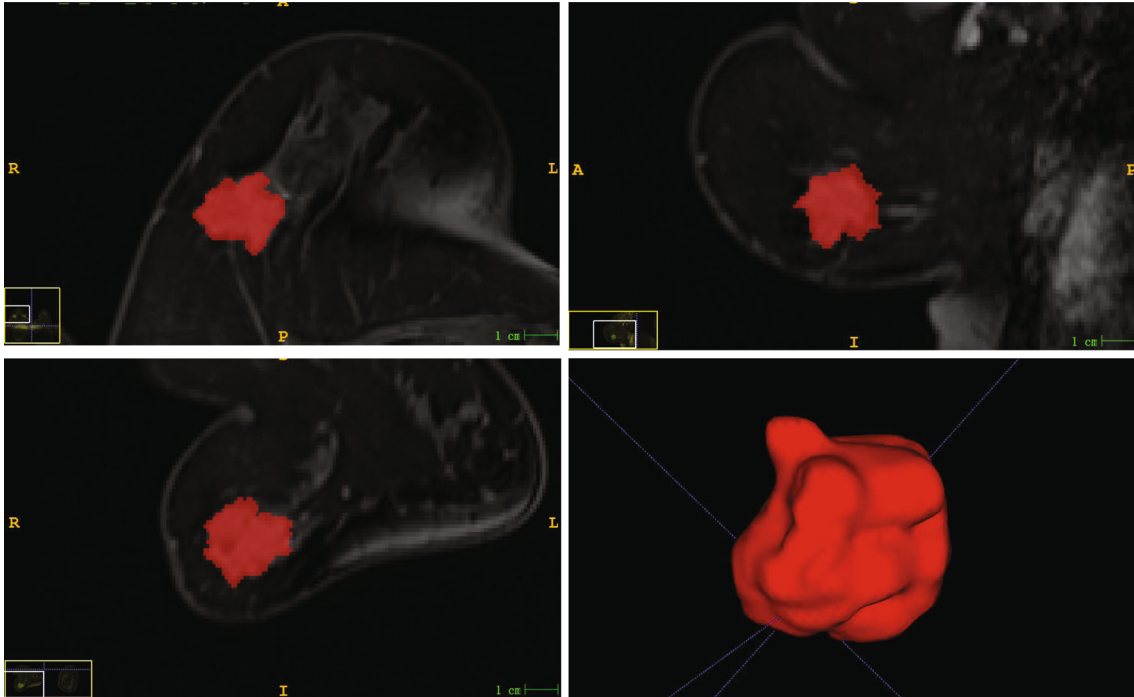


FIGURE 2: 3D-ROI segmentation method based on the lesion itself. Shows a case of invasive ductal carcinoma of the right breast (stage II).

inputting to the CNN, the pixel values of the image data are normalized to 0-1 range. The formula is as follows:

$$x = \frac{X - X_{\min}}{X_{\max} - X_{\min}}, \quad (1)$$

where  $x$  represents the normalized image pixel value,  $X$  represents the original image pixel value, and  $X_{\max}$  and  $X_{\min}$  represent the maximum pixel value and the minimum pixel value of the minimum bounding cube of all lesions, respectively.

In this study, a total of 207 masses were obtained, of which 106 were malignant and 101 were benign. The data



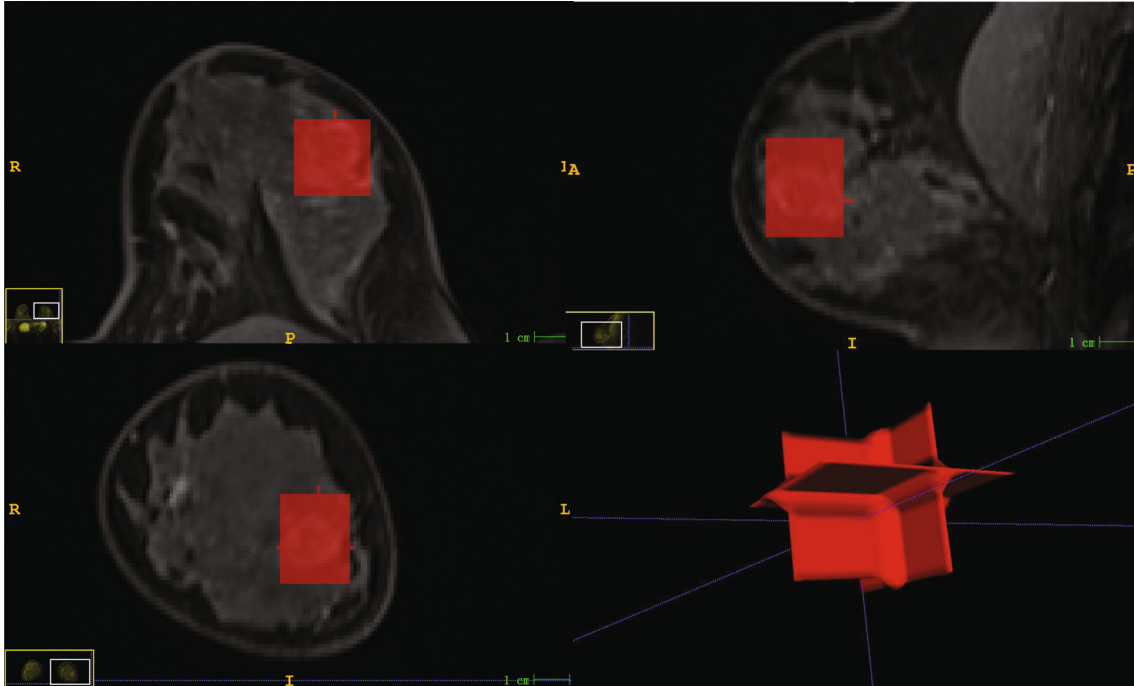


FIGURE 3: 3D-ROI segmentation method based on the minimum bounding cube at the edge of the lesion. Segmentation based on the minimum bounding box of the mass edge, showing a case of fibroadenoma of the left breast.

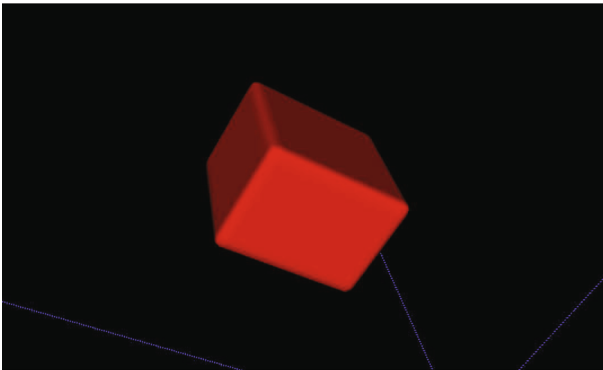


FIGURE 4: 3D-ROI segmentation method based on the minimum bounding cube at the edge of the lesion. The cubic model of segmentation based on the minimum bounding cube of the mass edge.

set was divided into the training set ( $n = 145$ ), the validation set ( $n = 22$ ), and the test set ( $n = 40$ ) at the ratio of 7:1:2. To avoid the selection bias of the benign and malignant lesions of the test set samples, the test set was generated by random sampling of 20 malignant lesions and 20 benign lesions. The remaining samples were randomly divided into the training set and the validation set in the ratio of 7:1. Fivefold cross-validation was performed.

**2.4. Model Establishment.** The model was established based on the blueprint of the classic residual network ResNet50 [19], retaining its residual module but changing the convolution module to a 3D convolution module. At the same time,

an attention mechanism was added to transform the attention mechanism module, which only fit the 2D image convolution module, into 3D-Convolutional Block Attention Module (CBAM) to adapt to 3D-MRI, as shown in Figure 5. CBAM includes a channel attention module and a spatial attention module, which together can solve the question of which channel and which position characteristics play decisive roles in final prediction [18]. Input module, residual module, channel attention module, downsampling module, and fully connected module constitute the main modules of the network. Among them, the residual module was mainly used to extract features, the CBAM module was mainly used to give higher weight to key features, and the downsampling module was used to reduce the size of the feature map and to increase the number of channels in the feature map. Blocks are used (Figure 5) to reflect the size change of the feature map. After the last CBAM, the algorithm stretches the output high-dimensional features into a one-dimensional vector and connects 2 fully connected slices. Lesion classification network parameters are shown in Table 1. The network uses cross-entropy cost function as the loss function and stochastic gradient descent (SGD) whose weight decay is 0.0001 and momentum is 0.9 as the optimizer. The batch size is 16. Dynamic learning rate strategy is taken during the train process. The initial learning rate is 0.1, which is considered as a big number, halved every 25 epochs of iterations. Before finally setting two output results (P1, P2), which, respectively, represent the probability of benign and malignant lesions. The lesion is classified as benign if  $P1 > P2$ . Otherwise, the lesion is classified as malignant.

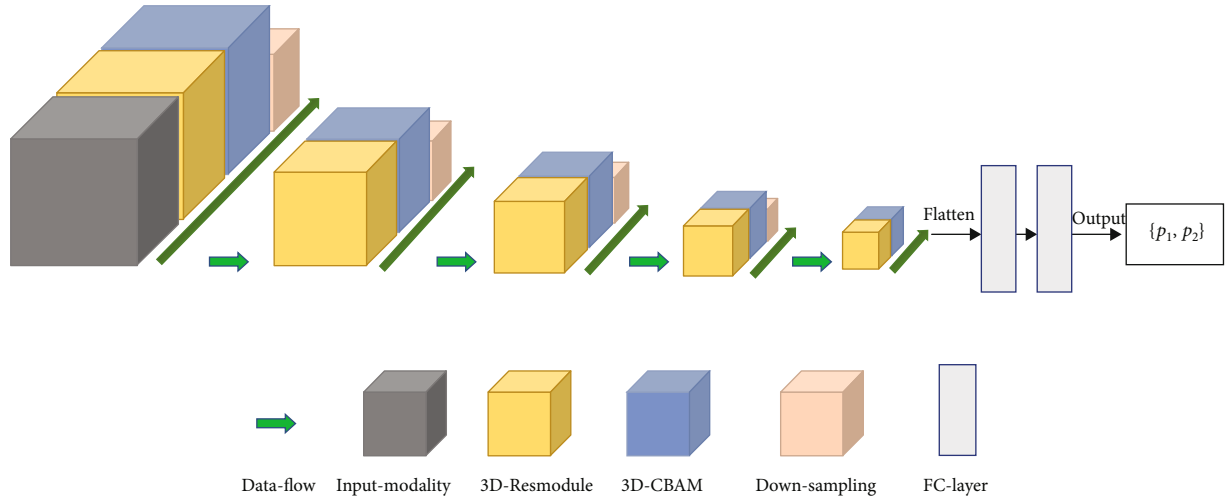


FIGURE 5: Structure of monomodal network.

TABLE 1: Tumor classification network parameters.

Layer_name	Input_size	Detailed_parameters	Output_size
conv	$1 * 64 * 64 * 64$	Kernel = 3, stride = 1, padding = 1	$8 * 64 * 64 * 64$
res_conv * 2	$8 * 64 * 64 * 64$	Kernel = 3, stride = 1, padding = 1	$8 * 64 * 64 * 64$
3D_CBAM	$8 * 64 * 64 * 64$	$1 * 1 * 1$ convolution	$8 * 64 * 64 * 64$
conv	$8 * 64 * 64 * 64$	Kernel = 2, stride = 2, padding = 0	$16 * 32 * 32 * 32$
res_conv * 4	$16 * 32 * 32 * 32$	Kernel = 3, stride = 1, padding = 1	$16 * 32 * 32 * 32$
3D_CBAM	$16 * 32 * 32 * 32$	$1 * 1 * 1$ convolution	$16 * 32 * 32 * 32$
conv	$16 * 32 * 32 * 32$	Kernel = 2, stride = 2, padding = 0	$32 * 16 * 16 * 16$
res_conv * 4	$32 * 16 * 16 * 16$	Kernel = 3, stride = 1, padding = 1	$32 * 16 * 16 * 16$
3D_CBAM	$32 * 16 * 16 * 16$	$1 * 1 * 1$ convolution	$32 * 16 * 16 * 16$
conv	$32 * 16 * 16 * 16$	Kernel = 2, stride = 2, padding = 0	$64 * 8 * 8 * 8$
res_conv * 4	$64 * 8 * 8 * 8$	Kernel = 3, stride = 1, padding = 1	$64 * 8 * 8 * 8$
3D_CBAM	$64 * 8 * 8 * 8$	$1 * 1 * 1$ convolution	$64 * 8 * 8 * 8$
conv	$64 * 8 * 8 * 8$	Kernel = 2, stride = 2, padding = 0	$128 * 4 * 4 * 4$
res_conv * 4	$128 * 4 * 4 * 4$	Kernel = 3, stride = 1, padding = 1	$128 * 4 * 4 * 4$
3D_CBAM	$128 * 4 * 4 * 4$	$1 * 1 * 1$ convolution	$128 * 4 * 4 * 4$
Flatten	$128 * 4 * 4 * 4$		8192
Full connection	8192		2048
Full connection	2048		512
Softmax	512		2

“res\_conv” is a residual convolution block which contains shortcut connection, and “res\_conv \* N” means the block has  $N$  convolution blocks that share the same parameters. 3D\_CBAM uses  $1 \times 1 \times 1$  convolutions to adjust the channel numbers of the current feature map.

**2.5. Model Evaluation.** The two DL feature models based on the lesion itself and the minimum boundary cube of lesion in the diagnosis of benign and malignant breast lesions were compared. Model performance was evaluated using the area under the ROC curve (AUC), accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), precision, and recall rate. The AUC of the two DL models was compared by the DeLong test.

### 3. Results

**3.1. Patient Characteristics.** A total of 296 patients with breast lesions were investigated in the study. The following cases were excluded: 3 patients with unqualified images due to the presence of motion artifacts, 12 patients with a history of biopsy or mastectomy before breast MRI examination, 53 patients with a single focal lesion or multiple focal

TABLE 2: Age and pathological information of patients with benign and malignant breast tumors.

	Benign patients ( $n = 98$ )	Malignant patients ( $n = 105$ )
Age (year)	$41.0 \pm 10.6$	$55.5 \pm 11.3$
Pathology (207 masses)	Fibrocystic adenosis ( $n = 18$ )	Invasive ductal carcinoma ( $n = 81$ )
	Fibroadenoma ( $n = 71$ )	Mucinous adenocarcinoma of the breast ( $n = 7$ )
	Intraductal papilloma ( $n = 4$ )	Invasive lobular carcinoma ( $n = 3$ )
	Intraductal papillomatosis ( $n = 4$ )	Intraductal carcinoma or tubular carcinoma ( $n = 8$ )
	Benign phyllodes tumor ( $n = 4$ )	Solid papillary carcinoma ( $n = 2$ )
		Malignant phyllodes tumor ( $n = 1$ )
		Small lymphocytic lymphoma ( $n = 1$ )
		Medullary carcinoma ( $n = 3$ )

lesions in the breast, 14 patients with incomplete examination or perfusion scan breast MRI, and 11 patients with breast lesions combined with nonmass enhancement lesions. Eventually, 203 patients were enrolled for analyses (Table 1). The patients were 17-86 years old with an average age of  $48.5 \pm 13.1$  years old. Among them, there was only one male patient, aged 54 years. There were 105 patients with malignant lesions with an average age of  $55.5 \pm 11.3$  years and 98 patients with benign lesions with an average age of  $41.0 \pm 10.6$  years old. A total of 207 masses were included in the study (Table 2).

### 3.2. Model Characteristics

*3.2.1. Comparison of the Two Deep Learning Feature Model Based on the Lesion Itself and the Minimum Boundary Cube of Lesion in the Diagnosis of Benign and Malignant Breast Lesions.* Using first postcontrast images of DCE-MRI T1WI sequence, the two deep learning feature model based on the lesion itself (model 1) and the minimum boundary cube of lesion (model 2) in the diagnosis of benign and malignant breast lesions were compared. The means are shown in Table 3. In comparison, the model 1 analysis achieved mean AUC of 0.799, accuracy of 74.63%, sensitivity of 83.65%, specificity of 65.35%, NPV of 71.31%, PPV of 79.52%, and recall rate of 83.65% and the model 2 analysis achieved an average AUC of 0.827, accuracy of 78.54%, sensitivity of 78.85%, specificity and PPV of 78.22%, NPV and recall rate of 78.85%. There was no statistical difference in AUC based on the lesion itself model and the minimum bounding cube model ( $Z = 0.771$ ,  $p = 0.4408$ ). The minimum bounding cube based on the edge of the lesion showed higher accuracy, specificity, and lower recall rate in identifying benign and malignant lesions.

## 4. Discussion

Deep learning in convolutional neural networks (CNNs) is usually based on manually or semiautomatically segmented tags to learn to recognize image features. Because breast MRI is different from MRI for abdomen and lung lesions, its position is fixed in a special breast coil and is less affected by breathing movement, leading to relatively higher reproducibility of the segmentation method for breast lesions.

TABLE 3: Comparison of two 3D-ROI segmentation methods based on the lesion itself and the minimum bounding cube at the edge of the lesion.

Evaluation index	Lesion itself	Minimum bounding cube
AUC	0.799	0.827
Accuracy (%)	74.63	78.54
Sensitivity (%)	83.65	78.85
Specificity (%)	65.35	78.22
Negative prediction (%)	71.31	78.85
Positive prediction (%)	79.52	78.22
Precision (%)	71.31	78.85
Recall (%)	83.65	78.85
DeLong test	$Z = 0.771$ , $p = 0.4408$	

However, the segmentation methods are quite different. Previous studies have mostly extracted the two-dimensional features of the lesion (2D-ROI) [20], selected the largest slice of the lesion or the most obvious slice of lesion enhancement [21], and segmented along the edge of the lesion. 2D-ROI can only represent the information covered by the current area and cannot reflect all the information of the lesion. Therefore, this will definitely affect the reliability of DL models. The use of 3D-ROI is helpful to observe the lesion's overall morphology, leading to more accurate and comprehensive reflection of the characteristics of the lesion [22]. And more weight is given to the hemodynamic characteristics of the relevant lesion in the model based on the usual imaging physicians' reading habits and the advantages of early enhanced MRI.

*4.1. The Efficacy of a Deep Learning Model Based on the Minimum Bounding Cube of the Lesion in Breast Lesion Classification.* This study used two different segmentation methods for 3D-ROI of the lesion: one was based on the lesion itself, and the other one was based on the minimum bounding cube of the lesion edge. These two different segmentation methods were compared for their impact on the accuracy of the DL model. Our results revealed that the DL model based on the minimum bounding cube of the lesion edge is more accurate, with a mean AUC value of about

0.827. The reason may be that the minimum bounding cube based on the lesion edge not only contains the internal information of the lesion but also includes some tissues surrounding the lesion.

Zhou et al. [23] applied 5 different input boxes (tumor alone, the smallest bounding box, and 1.2, 1.5, and 2.0 time box) in deep learning and showed that the performance of diagnosis gradually decreases as the bounding box increases. The per-lesion diagnostic accuracy was the highest when using the smallest bounding box (89%), but the tumor ROI on all slices were automatically segmented on contrast-enhanced maps by using the fuzzy-C-means (FCM) clustering algorithm with 3D connected-component labeling. This study used manually segmented images as a standard for comparison, which may be more accurate. And the minimum bounding cube based on the tumor edge did not expand the box size but instead used 3D-CBAM to increase the weight of key information, in order to prevent the box containing too much information from normal tissue that dilutes the effective information in the overall box or reduces the resolution of the effective information of the image imported into the neural network.

The DL model that is based on the minimum bounding cube of postcontrast images of DCE-MRI T1WI sequence showed superiority in the test set, a mean specificity of 78.22%, which are better than those of the DL model that is based on just the lesion itself. The reason may be that the microenvironment around the tumor plays a critical role in tumor growth and aggressive tissue behavior [24, 25]. 3D-CBAM was to give higher weight to those key features. The area around the tumor contains much valuable and hidden information about the disease, including survival predictors for vascular activity and lymphangiogenesis and the infiltration of lymphatic and blood vessels around the tumor, and immune response signals around the tumor for interstitial response and lymphocyte infiltration around the tumor [26]. As we have shown in a previous study [27], the peritumoral edema on T2WI images is better and appears as T2WI hyperintensity around the tumor. This sign is combined with the T2WI signal, leading to significantly increased sensitivity and specificity for the differential diagnosis of benign and malignant breast tumors, and there is a positive correlation between peritumoral edema and Ki-67 expression. These results demonstrate the importance of the tissue surrounding the tumor. However, related studies are still limited at present; thus, the information about surrounding tissues has not been captured by the artificial intelligence learning technology. Braman et al. [26] collected a total of 117 patients and extracted omics features after marking the breast tumors and surrounding areas (2.5-5 mm area around the tumor) using breast images from DCE-MRI-T1WI. Their results showed that the omics features of surrounding tissues helped to predict pCR and that combined use of tumor internal characteristics and peritumoral characteristics led to better prediction accuracy, which as a whole may help guide the personalized treatment of locally advanced breast cancer. This indicates that extracting the information contained in the tissue around the tumor has a high clinical application value.

*4.2. The Diagnostic Efficacy of the Deep Learning Model Based on First Postcontrast Images of DCE-MRI T1WI Sequence in Benign and Malignant Breast Lesions.* The deep learning model that is based on the minimum bounding cube of dynamic contrast postcontrast images has high specificity in the classification of benign and malignant breast lesions. We speculate that this may be related to the early hemodynamic information of the lesion, as shown in a previous study of ours that DCE-MRI can not only reveal tumor's morphological changes but also reflect its microvascular perfusion, angiogenesis, grades, and malignancy for evaluating the effect of tumor treatment and prognosis. The degree of early enhancement reflects the abundance of blood vessels and blood perfusion of the disease [28]. Malignant lesions grow fast, have multiple large blood vessels, are immature, and have a large number of arteriovenous anastomoses.

In addition to the high accuracy in diagnostic performance of the minimum bounding cube based on the edge of the lesion, we also found that the method is relatively simple and easy to use, as it only needs to find the largest level of the three dimensions of the image through image processing software. At this level, the minimum rectangle that can cover the outermost edge of the lesion is used, and finally, the minimum bounding cube containing the lesion is generated by the computer traversal method. However, the 3D-ROI based on the lesion itself needs to be delineated slice by slice and along the edge. For nonenhancement sequence images, sometimes, the edge of the lesion is unclear, leading to the lack of local edge information of the lesion.

## 5. Conclusion

In summary, based on the segmentation method of the minimum bounding cube at the edge of the lesion, postcontrast images of DCE-MRI T1WI sequence were extracted, and a DL model was established. This model can combine the information inside the lesion and that of containing peritumoral area to improve the diagnostic efficacy for both benign and malignant breast lesions. Using the DL approach with a 3D attention mechanism based on ResNet50 to identify benign and malignant BI-RADS 4 lesions was feasible.

## 6. Limitations of This Study

This study was a small-sample single-center study, and the results obtained in this study need to be confirmed by future large-sample multicenter investigations. Only mass lesions were included in the study; thus, whether the segmentation method is equally applicable to nonmass lesions remains to be tested. The inclusion/exclusion criteria are quite stringent and exclude many of the lesions which a radiologist reading breast MRI will routinely come across. The study only used first postcontrast images of DCE-MRI T1WI for segmentation by the minimum bounding cube of the lesion, which needs to be examined to see if it fits other sequences of image segmentation. Another limitation is that this study only compared two lesion segmentation methods; thus, future

investigation is needed to test whether other ROIs containing peritumoral area may be better.

## Data Availability

All data generated or analyzed during this study are available from the corresponding author Wei Xing upon reasonable request.

## Ethical Approval

The retrospective study was approved by the Ethical Review Board of Nantong First People's Hospital (No. 2020KY236) and was conducted according to the Declaration of Helsinki principles.

## Consent

All patients signed informed consent.

## Conflicts of Interest

The authors declare that they have no conflict of interest.

## Authors' Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis, and interpretation or in all these areas; took part in drafting, revising, or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work. Meihong Sheng and Weixia Tang are co-first authors.

## Acknowledgments

This project was funded by the Jiangsu Province Maternal and Child Health Research Project (F202037), the "Six One" Research Funding Program for high-level health talents in Jiangsu Province (LGY2018036 and LGY2020048 in 2018 and 2020), and the Jiangsu Province Science and Technology Project (BE2018646).

## References

- [1] H. Sung, J. Ferlay, R. L. Siegel et al., "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a Cancer Journal for Clinicians*, vol. 71, no. 3, pp. 209–249, 2021.
- [2] P. D. P. Pharoah, B. Sewell, D. Fitzsimmons, H. S. Bennett, and N. Pashayan, "Cost effectiveness of the NHS breast screening programme: life table model," *BMJ*, vol. 346, p. f2618, 2013.
- [3] B. Lauby-Secretan, C. Scoccianti, D. Loomis et al., "Breast-cancer screening —viewpoint of the IARC working group," *The New England Journal of Medicine*, vol. 372, no. 24, pp. 2353–2358, 2015.
- [4] M. G. Marmot, D. G. Altman, D. A. Cameron, J. A. Dewar, S. G. Thompson, and M. Wilcox, "The benefits and harms of breast cancer screening: an independent review," *Lancet*, vol. 380, pp. 1778–1786, 2012.
- [5] H. I. Greenwood, L. J. Wilmes, T. Kelil, and B. N. Joe, "Role of breast MRI in the evaluation and detection of DCIS: opportunities and challenges," *Journal of Magnetic Resonance Imaging*, vol. 52, no. 3, pp. 697–709, 2020.
- [6] E. A. Morris, C. E. Comstock, and C. H. Lee, *ACR BI-RADS Magnetic Resonance Imaging*, in: *American College of Radiology, BI-RADS Committee, Editor. ACR BI-RADS Atlas: Breast Imaging Reporting and Data System*, American College of Radiology, Reston, 5th edition, 2013.
- [7] R. M. Strigel, E. S. Burnside, M. Elezaby et al., "Utility of BI-RADS assessment category 4 subdivisions for screening breast MRI," *AJR. American Journal of Roentgenology*, vol. 2086, pp. 1392–1399, 2017.
- [8] J. R. Maltez de Almeida, A. B. Gomes, T. P. Barros, P. E. Fahel, and M. de Seixas Rocha, "Subcategorization of suspicious breast lesions (BI-RADS category 4) according to MRI criteria: role of dynamic contrast-enhanced and diffusion-weighted imaging," *AJR. American Journal of Roentgenology*, vol. 205, no. 1, pp. 222–231, 2015.
- [9] N. Houssami, S. Ciatto, I. Ellis, and D. Ambrogetti, "Underestimation of malignancy of breast core-needle biopsy: concepts and precise overall and category-specific estimates," *Cancer*, vol. 109, no. 3, pp. 487–495, 2007.
- [10] K. Doi, "Computer-aided diagnosis in medical imaging: historical review, current status and future potential," *Computerized Medical Imaging and Graphics*, vol. 31, no. 4-5, pp. 198–211, 2007.
- [11] Y. Zhang, S. Wang, K. Xia, Y. Jiang, and P. Qian, "Alzheimer's disease multiclass diagnosis via multimodal neuroimaging embedding feature selection and fusion," *Information Fusion*, vol. 66, pp. 170–183, 2021.
- [12] Y. Zhang, Z. Zhou, H. Bai, W. Liu, and L. Wang, "Seizure classification from EEG signals using an online selective transfer TSK fuzzy classifier with joint distribution adaption and manifold regularization," *Frontiers in Neuroscience*, vol. 14, p. 496, 2020.
- [13] Y. Zhang, F. Chung, and S. Wang, "Clustering by transmission learning from data density to label manifold with statistical diffusion," *Knowl Based Syst*, vol. 193, p. 105330, 2020.
- [14] A. Hosny, C. Parmar, J. Quackenbush, L. H. Schwartz, and H. J. W. L. Aerts, "Artificial intelligence in radiology," *Nature Reviews. Cancer*, vol. 18, no. 8, pp. 500–510, 2018.
- [15] M. Khalili and R. J. Wong, "Underserved does not mean undeserved: unfurling the HCV care in the safety net," *Digestive Diseases and Sciences*, vol. 63, no. 12, pp. 3250–3252, 2018.
- [16] C. Vasile, A. L. Udriștoiu, A. E. Ghenea et al., "Intelligent diagnosis of thyroid ultrasound imaging using an ensemble of deep learning methods," *Medicina (Kaunas)*, vol. 57, no. 4, p. 395, 2021.
- [17] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097–1105, 2012.
- [18] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: convolutional block attention module," *European Conference on Computer Vision (ECCV)*, pp. 3–19, 2018.
- [19] S. M. Anwar, M. Majid, A. Qayyum, M. Awais, M. Alnowami, and M. K. Khan, "Medical image analysis using convolutional

- neural networks: a review,” *Journal of Medical Systems*, vol. 42, no. 11, p. 226, 2018.
- [20] P. Herent, B. Schmauch, P. Jehanno et al., “Detection and characterization of MRI breast lesions using deep learning,” *Diagnostic and Interventional Imaging*, vol. 100, no. 4, pp. 219–225, 2019.
- [21] K. Holli-Helenius, A. Salminen, I. Rinta-Kiikka et al., “MRI texture analysis in differentiating luminal A and luminal B breast cancer molecular subtypes - a feasibility study,” *BMC Medical Imaging*, vol. 17, no. 1, p. 69, 2017.
- [22] C. Song, W. S. Zhu, S. Y. Shi et al., “Differentiation between benign and malignant non-mass enhancement lesions using volumetric quantitative dynamic contrast-enhanced MR imaging,” *Radiology Practice*, vol. 35, no. 2, pp. 190–196, 2020.
- [23] J. Zhou, Y. Zhang, K. T. Chang et al., “Diagnosis of benign and malignant breast lesions on DCE-MRI by using radiomics and deep learning with consideration of peritumor tissue,” *Journal of Magnetic Resonance Imaging*, vol. 51, no. 3, pp. 798–809, 2020.
- [24] Y. Kim, M. A. Stolarska, and H. G. Othmer, “The role of the microenvironment in tumor growth and invasion,” *Progress in Biophysics and Molecular Biology*, vol. 106, no. 2, pp. 353–379, 2011.
- [25] J. S. Wu, S. R. Sheng, X. H. Liang, and Y. L. Tang, “The role of tumor microenvironment in collective tumor cell invasion,” *Future Oncology*, vol. 13, no. 11, pp. 991–1002, 2017.
- [26] N. M. Braman, M. Etesami, P. Prasanna et al., “Intratumoral and peritumoral radiomics for the pretreatment prediction of pathological complete response to neoadjuvant chemotherapy based on breast DCE-MRI,” *Breast Cancer Research*, vol. 19, no. 1, p. 57, 2017.
- [27] W. X. Tang, M. H. Sheng, S. C. Gong, H. T. Chen, J. B. Ge, and Y. Zhu, “Diagnostic value of T2WI relative signal intensity ratio combined with peritumoral edema in benign and malignant breast lesions,” *Journal of Clinical Radiology*, vol. 39, no. 12, pp. 2411–2414, 2020.
- [28] M. H. Sheng, W. X. Tang, Y. H. Lu et al., “Value of early-phase enhancement ratio combined with peripheral vascular diameter in the differential diagnosis of benign and malignant breast lesions under dynamic contrast enhanced MRI,” *Chinese Journal of Radiology*, vol. 50, no. 5, pp. 324–328, 2016.

## Research Article

# Analysis of Japanese Expressions and Semantics Based on Link Sequence Classification

Yanyan Shi<sup>1</sup> and Yuting Liang<sup>2</sup>

<sup>1</sup>*School of Foreign Languages, Harbin University of Commerce, Harbin Heilongjiang 150028, China*

<sup>2</sup>*Department of Foreign Languages, East University of Heilongjiang, Harbin Heilongjiang 150066, China*

Correspondence should be addressed to Yanyan Shi; [shiraohekaw@163.com](mailto:shiraohekaw@163.com)

Received 17 July 2021; Revised 8 August 2021; Accepted 17 August 2021; Published 6 September 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Yanyan Shi and Yuting Liang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Based on locness corpus, this paper uses WordSmith 6.0, SPSS 24, and other software to explore the use of temporal connectives in Japanese writing by Chinese Japanese learners. This paper proposes a method of tense classification based on the Japanese dependency structure. This method analyzes the results of the syntactic analysis of Japanese dependence and combines the tense characteristics of the target language to extract tense-related information and construct a maximum entropy tense classification model. The model can effectively identify the tense, and its classification accuracy shows the effectiveness of the classification method. This paper proposes a temporal feature extraction algorithm oriented to the hierarchical phrase expression model. The end-to-end speech recognition system has become the development trend of large-scale continuous speech recognition because of its simplicity and efficiency. In this paper, the end-to-end technology based on link timing classification is applied to Japanese speech recognition. Taking into account the characteristics of Japanese hiragana, katakana, and Japanese kanji writing forms, through experiments on the Japanese data set, different suggestions are explored. The final effect is better than mainstream speech recognition systems based on hidden Markov models and two-way long and short-term memory networks. This algorithm can extract the temporal characteristics of rules that meet certain conditions while extracting expression rules. These tense characteristics can guide the selection of rules in the expression process, make the expression results more in line with linguistic knowledge, and ensure the choice of relevant vocabulary and the structural ordering of the language. Through the analysis of time series and static information, we combine the time and space dimensions of the network structure. Using connectionist temporal classification (CTC) technology, an end-to-end speech recognition method for pronunciation error detection and diagnosis tasks is established. This method does not require phonemic information nor does it require forced alignment. The extended initials and finals are the error primitives, and 64 types of errors are designed. The experimental results show that the method can effectively detect the wrong pronunciation, the detection accuracy rate is 87.07%, the false rejection rate is 7.83%, and the error rate is 87.07%. The acceptance rate is 25.97%. This method uses network information more comprehensively than traditional methods, and the model is more effective. After detailed experiments, this article evaluates the prediction effect of this method and previous methods on the data set. This method improves the prediction accuracy by about 15% and achieves the expected goal of the work in this paper.

## 1. Introduction

Statistical language expression is one of the challenging frontier topics in the field of natural language processing, which has a wide range of application value and important commercial application prospects [1]. In recent years, statistical language expression technology has developed rapidly, and

a series of impressive results have been achieved. However, in practical applications, how to effectively use linguistic knowledge in statistical language expression models to improve the quality of expression is still a research hotspot [2]. At present, in the statistical machine expression, the research on tense is mainly limited to the aspect of tense recognition, and there are few studies on the expression of tense

[3]. Temporal information is important linguistic information, so the tense problem studied in this paper is transformed into a problem of incorporating tense and other linguistic knowledge into statistical expressions [4]. Driven by many applications, the research on link prediction has achieved fruitful results. At present, a method based on node similarity is widely used, and the possibility of link generation is predicted by the size of the similarity score [5]. In the static method, the network changes over time are ignored. If only the network diagram under the most recent time snapshot is used, when the network changes frequently, the prediction effect will drop sharply [6]. With the development of the Internet, there are more and more extensive scenarios where links occur repeatedly, and the evolution of networks is becoming more and more common. Static link prediction methods are far from being able to adapt to the needs of the new situation. Therefore, in recent years, information has gradually gained attention [7].

In recent years, with the rapid development of the Internet at home and abroad and the integration of the world economic market, the amount of network data information has increased sharply, international exchanges have become more frequent, and the language expression market is broad [8]. Language expression, as an important way to overcome language communication barriers, has a profound impact on the promotion of political, economic, cultural, and military exchanges between countries [9]. Language expression is a powerful guarantee for active and healthy exchanges and dialogues between the two countries. However, traditional manual expression is inefficient and costly. For the huge number and increasing number of expression tasks, it can no longer meet the needs of society and the market [10]. Language expression is the automatic expression of text, which is an important content in text processing, and most applications in text processing need to reason and filter according to the temporal relationship of text, such as time extraction and automatic summarization, and tense can provide them important clues, so tense plays an indispensable role in these applications [11]. Similarly, the correct expression of the tense in the text can convey the information of the source language as accurately as possible, but different languages have greater differences in the expression of tense, which is a great challenge to the expression of the tense in terms of language [12]. At the technical level, the existing language expression methods are still mainly limited to the use of rules to solve tense problems. These methods are inefficient and costly. Even statistical language expression methods cannot solve language well [13].

This article takes Japanese as the research object and studies the expression of tense from the perspectives of Japanese-Chinese. Japanese belongs to the cohesive language family, and its tense is determined by the deformation of the predicate ending, and the changes of the predicate ending are various. There are similar endings in different simultaneous expressions, which leads to the low accuracy of the tense expression of statistical expressions. In response to the above problems, this article proposes a statistical expression method that incorporates tense characteristics. The difference between a recursive network and a feedforward

network lies in this feedback loop that constantly uses its own output at the last moment as input. The purpose of adding memory to the neural network is the sequence itself contains information, and the recursive network can use this information to complete tasks that the feedforward network cannot complete. This sequence information is stored in the hidden state of the recursive network and is continuously passed to the previous layer, spanning many time steps, affecting the processing of each new sample. Human memory will continue to cycle invisible in the body, affecting our behavior without showing a complete appearance, and information will also circulate in the hidden state of the recursive network. This method uses a deep control gating function to connect multilayer LSTM units and introduces linear correlation between the upper and lower layers in the recurrent neural network, which can build a deeper voice model. At the same time, we use the training criteria of linking time series classification for model training. We build an end-to-end speech recognition system to solve the hidden Markov model's need to force the alignment of labels and sequences and use CTC training criteria to realize an end-to-end speech system, by combining with the traditional LSTM-CTC model. The model is compared to verify the effectiveness of the deep LSTM neural network [14, 15] in speech recognition.

## 2. Related Work

Rule-based expression technology is highly dependent on humans. It mainly relies on linguists to summarize rules and manually compile the rules. The workload is large, and problems such as rule conflicts are prone to occur, and it is difficult to cope with large-scale expression tasks. The corpus-based method pays more attention to automatically obtaining rules from a large-scale corpus, which has the advantages of language independence and automatic knowledge acquisition, which greatly improves the efficiency of expression. With the increase of the amount of network information, the rule-based method has become more and more limited, and the corpus-based method has more and more obvious advantages, has developed rapidly, and has achieved a series of remarkable results [16]. Allen et al. [17] proposed an expression model based on word alignment, which entered the development stage of language expression and marked the birth of modern statistical expression methods. McCune [18] proposed the phrase expression model, which greatly improved the expression effect. This phrase expression model used a logarithmic linear model and its weight tuning method, which has made a great contribution to the expression performance improvement of the phrase expression model. At the same time, due to its strong ease of use and other advantages, the phrase expression model has begun to attract wide attention from all walks of life. In practical applications, the online translation systems of large domestic and foreign Internet companies such as Baidu and Google all use this model as backend support, which has certain commercial value. However, this model uses phrases as the basic processing unit, and its ability to adjust order is limited, and the quality of



long sentences is not good, which needs further improvement. IZARD et al. [19] introduced the concept of generalized variables. A hierarchical phrase expression model was further proposed, which used hierarchical phrase rules to achieve expression, which made the entire expression process more hierarchical, to a certain extent, alleviates the problem of insufficient global ordering ability in the phrase expression model, and solved other problems of ordering. The method of the problem has a certain guiding effect. As an expression model based on formal grammar, this model introduces more effective information for the expression process. Similarity can be described by many methods. According to the type of information used, link prediction mainly includes similarity methods based on network topology and similarity methods based on node attributes. The network topology can be divided into local information and global information. Asahara and Matsumoto [20] also considered the number of neighbors in common neighbors. Later, these neighbor-based methods were also extended, not only considering whether there is a link between neighbors but also the number of links, that is, the addition based on the basic neighbor method. In addition to neighbor-based methods, global information was also often used for link prediction, such as path information between nodes. Another way to consider global information is the random walk model. This model can be considered a general expression based on the neighbor method and the path-based method, which can better reflect the network topology information. The similarity method based on node attributes is rarely used solely for link prediction. Because on the one hand, node attributes are often used in specific types of networks, and the processing is relatively complicated; on the other hand, node attributes may be subjective and sometimes not as reliable as the network topology. However, node attributes and network topology, respectively, reflect information from different aspects of the network, and combining the two can often achieve a better effect. In addition to node attributes and network topology, community information has recently been shown to be helpful for link prediction. At present, various types of information in networks such as local information, global information, and community information are often targeted at specific fields, and the relatively comprehensive and comprehensive similarity score model is not very satisfactory.

Since tenses have different and complex expressions in different languages, it is very difficult to maintain the same tense information from the source language to the target language in terms of expression. In the machine expression system based on intermediate language conversion [21], the tense information of the source language was first converted into language-independent abstract expressions. For example, in the study of Chinese-Japanese tense expression, Cheng et al. [22] proposed Lexical Conceptual Structures (LSC) based on two levels of knowledge expression are used to assist language expression, combining tense and lexical semantics, and through some rule transformations in the expression process, it can generate Japanese sentences with correct tenses for Chinese sentences expression. However, the above research is based on a language expression system based on intermediate language conversion. It is a rule-based

language expression. It is not only highly subjective but also highly dependent on language types, and it is difficult to expand. It is also significantly different from the existing SMT system. On this basis, others jointly build a temporal model with the context of the target end. First, they syntactically analyzed the bilingual parallel corpus and used the syntactic analysis result to automatically extract the temporal information contained in the sentence, and then according to the temporal continuity and text classification, a temporal model based on the classifier was proposed. In the SMT system, someone proposed a corpus-based method to study Japanese and Chinese tense expression, but mainly demonstrated that bilingual corpus contains rich tense information, which can help solve the problem of Japanese and Chinese tense expression. On this basis, we can consider integrating bilingual tense information to solve the problem of tense expression [23]. In the statistical language expression system, on the one hand, there is less research on the expression of tense. On the other hand, the expression research on the integration of linguistic knowledge is in its infancy. Various methods need to be further improved. The full use of linguistic knowledge in the research of machine translation has important value and also has important guiding significance for the research of tense in this article.

### 3. Japanese Language Expression and Semantic Model Construction Based on Link Timing Classification

*3.1. Link Timing Classification.* Time series has achieved good results in describing time information. There are two main ways to represent the network information of each time period in history as discrete time series diagrams and link predictions. Link timing classification (CTC) is mainly used to deal with timing classification tasks, especially when the alignment result between the input signal and the target label is unknown. Link timing classification technology can perform label prediction at any point in the entire input sequence, which solves the problem of forced alignment in traditional speech recognition. The criterion for neural network training by linking time series classification technology is called the CTC criterion. Figure 1 shows the hierarchical distribution of link timing classification.

The existing information of the static method is a graph. According to the link between nodes  $x$  and  $y$  and other nodes, the similarity scores of  $x$  and  $y$  are calculated to predict the current unlinked node pair  $(x, y)$  will be generated in the next time period possibility of linking.

$$X[t] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \dots & 0 \\ 0 & 0 & t \end{bmatrix}, \quad (1)$$

$$Y(x) = x(1) + 2 \times x(2) + \dots + t \times x(t). \quad (2)$$

One is the time series of the number of links between nodes, which only predicts the future link situation based on the past links between nodes, and achieves similar results

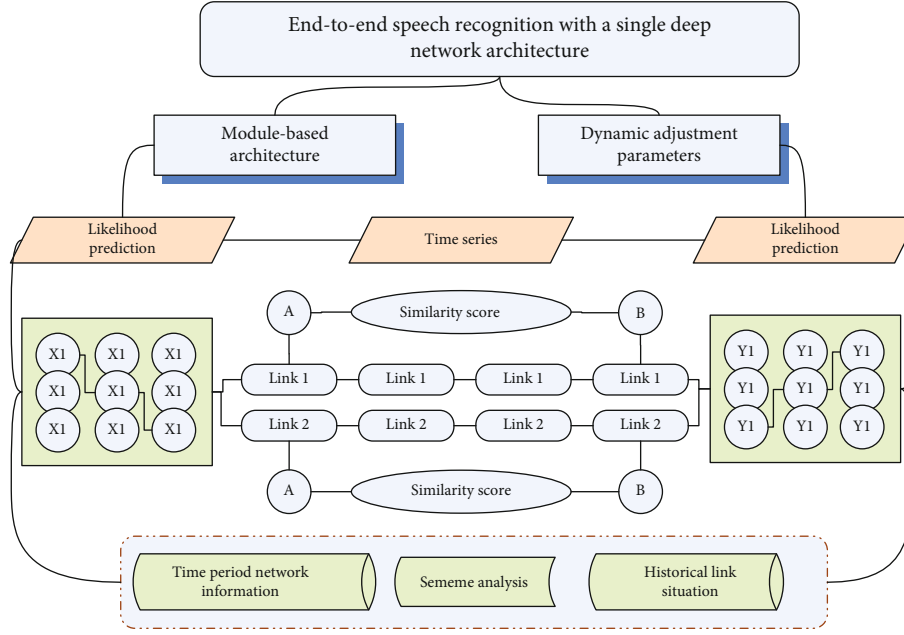


FIGURE 1: Hierarchical distribution of link timing classification.

to the static method. Combining it with the static method can further improve the prediction effect.

$$C(t, s) = \{t \in A : s(t) = t * (t - 1), t \rightarrow N\}. \quad (3)$$

For new links, because the time series of link times is lost, the hybrid model is downgraded to a static similarity method; in addition, the hybrid model multiplies the final static method prediction value with the time series prediction value, which makes it difficult to describe the network in each time period.

$$z = \begin{cases} i \times (x(t) + x(t - 1)), t > i, \\ i \times (x(t) - x(t - 1)), t < i, \end{cases} \quad (4)$$

$$H[x] = [x(1) \cdots x(t)] \times \begin{bmatrix} x(t) \\ \cdots \\ x(1) \end{bmatrix}. \quad (5)$$

The existing information of this method is from the first time period in history to the current time period  $t$ , a total of  $t$  pictures. According to  $(x, y)$  historical link situation (including the bold dashed line at time  $t$ ), we establish a unary time series model to predict the link situation in the next time period. The probability number of links is greater than 0 is the link probability.

$$p(t|x) - \prod_{t=1}^T H(x) \times (x(t) - \bar{x}) = 0, \quad (6)$$

$$\frac{\partial \ln p(z|x)}{\partial x} \times p(z|x) \times H(x) = 1. \quad (7)$$

However, because the model is too simple and fails to describe the relationship between the similarity score and the number of links, the changing laws of the two are different, and the results obtained by the mixed model are not as good as just using the similarity score time series.

**3.2. Temporal Feature Extraction.** Time series is a sequence of ordered data recorded in chronological order. We observe the time series in order to find the law of its development and change, and then, we select a suitable model to fit the observations to complete the prediction of the future value based on the model. The application of time series analysis and forecasting methods is very rich, such as forecasting the load of the power system and the price of a stock in the stock market.

Figure 2 shows the flow of sentence temporal feature extraction. The first part is the input layer, which accepts the acoustic characteristics of the input. As a branch of mathematical statistics, time series has its own set of analysis and prediction methods. In the unary time series model, in order to predict the future value, the historical value of the series itself is the research focus. Different from the unitary time series, in addition to its own influence, the changing laws of some series are also related to other series. The next is the batch normalization layer and the zero-filling layer. The reason for adding the zero-filling layer is to ensure batch processing with the same length; the second part is convolution, which contains 10 CNN layers, 5 maxpool layers, then the batch normalization layer, and the dropout layer. The function of the dropout layer is to prevent overfitting and improve generalization ability during training; the third part is the fully connected layer (dense layer), each neuron in the fully connected layer is connected with all neurons in the previous pooling layer, and the fully connected layer integrates the classification features of the convolution and

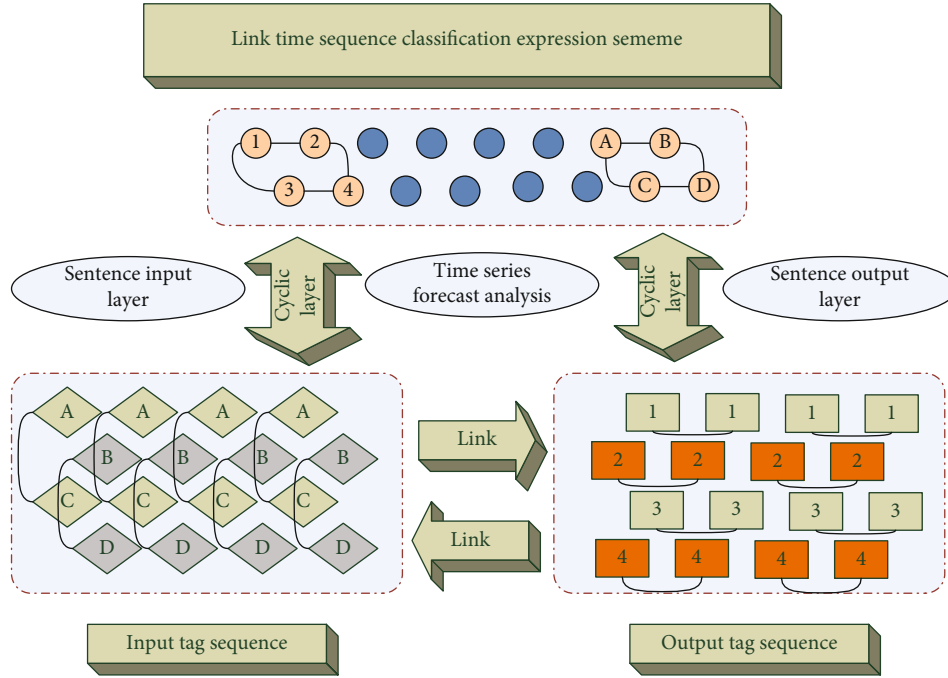


FIGURE 2: Sentence temporal feature extraction process.

pooling layers and distinguishes them. The activation function of each neuron uses the linear rectification function. The output value of the last layer is passed to the softmax logistic regression for classification. Finally, the CTC output layer is used to generate the predicted phoneme sequence. This method also tries to compare the similarity score calculated by the entire network between nodes and the actual occurrence between nodes. Combining the number of links, the hybrid model normalizes the similarity score of each time period and adds the number of links, which is used as the input of the time series.

**3.3. Phrases and Semantic Fusion.** The language model of the encoder-decoder structure [24] integration is only reflected in the prediction of the current moment by using the prediction of the previous moment during decoding. There are certain limitations and cannot make full use of linguistic information. Therefore, in order to explicitly introduce it in the decoding process in the language model, we further improve the performance of speech recognition. Based on the transformer structure, this article removes the decoder part and combines the encoder with the link timing classification as the end-to-end model used in this article. The main body of the entire model is composed of several identical coding layers stacked, and each coding layer is divided into two parts: multihead attention [25, 26] and feed-forward network. The main method of current speech recognition is to train the acoustic model by combining the recurrent neural network (RNN) and its variants with the hidden Markov model. The cyclic neural network uses the past information to input the output of the hidden layer at the last moment into the hidden layer at the current moment, retaining the previous information. As a time series, the speech signal has a strong context dependence, so the recur-

rent neural network is quickly applied to speech recognition. Theoretically, RNN can handle arbitrarily long sequences, but due to the disappearance of the gradient, the RNN cannot use the information at a longer time. A residual connection is added after each part, and then, layer normalization is performed. Figure 3 shows the three-dimensional histogram of the residuals in the coding layer of the language model. In addition, the model also includes a downsampling and upsampling module, a position encoding module, and an output layer. The output adopts the link timing classification criterion and predicts a label for each frame of speech to minimize the CTC loss function.

The “multihead attention” module is composed of several identical layers stacked, and each layer is a self-attention mechanism that uses scaled dot-product attention (scaled dot-product attention). Self-attention is a mechanism that uses the connection between different positions of the input sequence to calculate the input representation. Specifically, it has three inputs, queries, keys, and values, which can be understood as the speech feature after encoding. The output of query is obtained by the weighted summation of value, and the weight of each value is calculated by the design function of query and its related key. The “multihead” mechanism is used to combine multiple different layers. The self-attention representation of that is calculated, and  $h$  represents the number of “heads”. Figure 4 shows the fusion of linguistic expressions and sememe characters. Combining the characters in the word with spaces as separators to form character pairs, we combine word frequency and count all character pairs and their appearance frequency. We select the pair of characters with the highest frequency, remove the spaces in the middle, and merge them. The resulting character sequence is used as a new symbol to replace the original character pair.

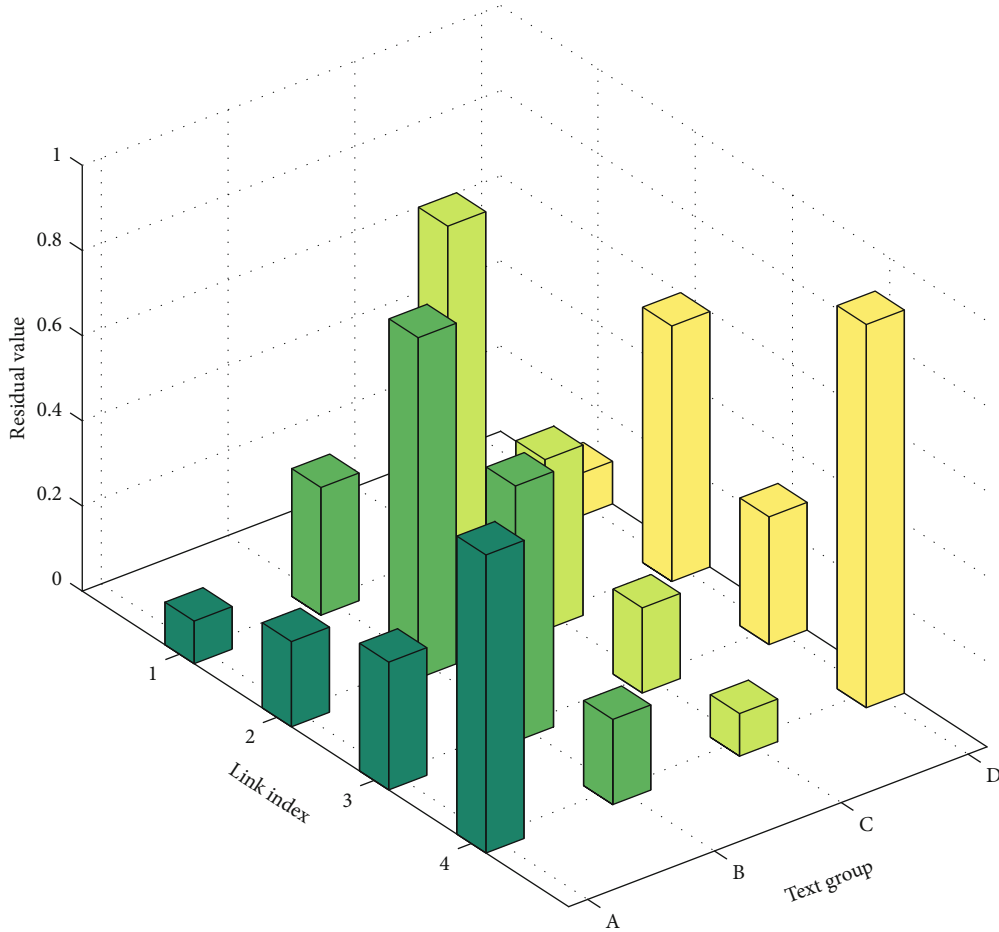


FIGURE 3: Three-dimensional histogram distribution of residuals in the coding layer of the language model.

Then, we repeat the two steps of counting the frequency of character pairs and merging the most frequent character pairs until the number of merging reaches the specified number. We output the merged character pairs according to the statistical frequency from high to low, then use the character pair list generated in the previous section. The corpus that needs to be processed is segmented. When segmenting is performed, the same word is used as the unit, and the characters are first segmented. Then, the character pairs are merged in the order of the frequency of appearance in the character pair list from high to low. Finally, it is retained that does not appear in units in the list of character pairs and individual characters not participating in the merging.

#### 4. Application and Analysis of Japanese Language Expression and Semantic Model Based on Link Timing Classification

*4.1. Link Timing Score Prediction.* In the LSTM-based end-to-end speech recognition system, we use a bidirectional long and short-term memory network to model the timing of speech features, and the output uses the link timing classification criterion to directly predict the label sequence. In this experiment, we use 3 hidden layers. They are combined with a

LSTM network with 1024 hidden nodes in each layer and 108-dimensional filterbank features for acoustic model training. The modeling unit uses ub-word units. The 3-gram language model is combined when decoding. Figure 5 shows the sentence level accuracy deviation box type figure. In recognition, in order to use the knowledge of Japanese linguistics to further improve the recognition performance, we combined the traditional language model when decoding and unified the dictionary, language model, and acoustic model to decode, and the system recognition performance was significantly improved. On the Japanese data set, using the algorithm recommended in this article, the recognition performance is significantly better than the current mainstream hybrid system based on the hidden Markov model and the end-to-end model based on the bidirectional long and short-term memory network. In order to reduce the memory occupied during model training and speed up the training, we first downsample the original speech feature frame and, then, encode it through the linear layer. At the same time, because the output adopts the CTC criterion, it is necessary to predict a label for each frame of speech, so the speech feature is performed before the output layer. One-step upsampling operation restores the length of the speech frame to the original length. In this way, while speeding up the calculation and reducing memory requirements, it

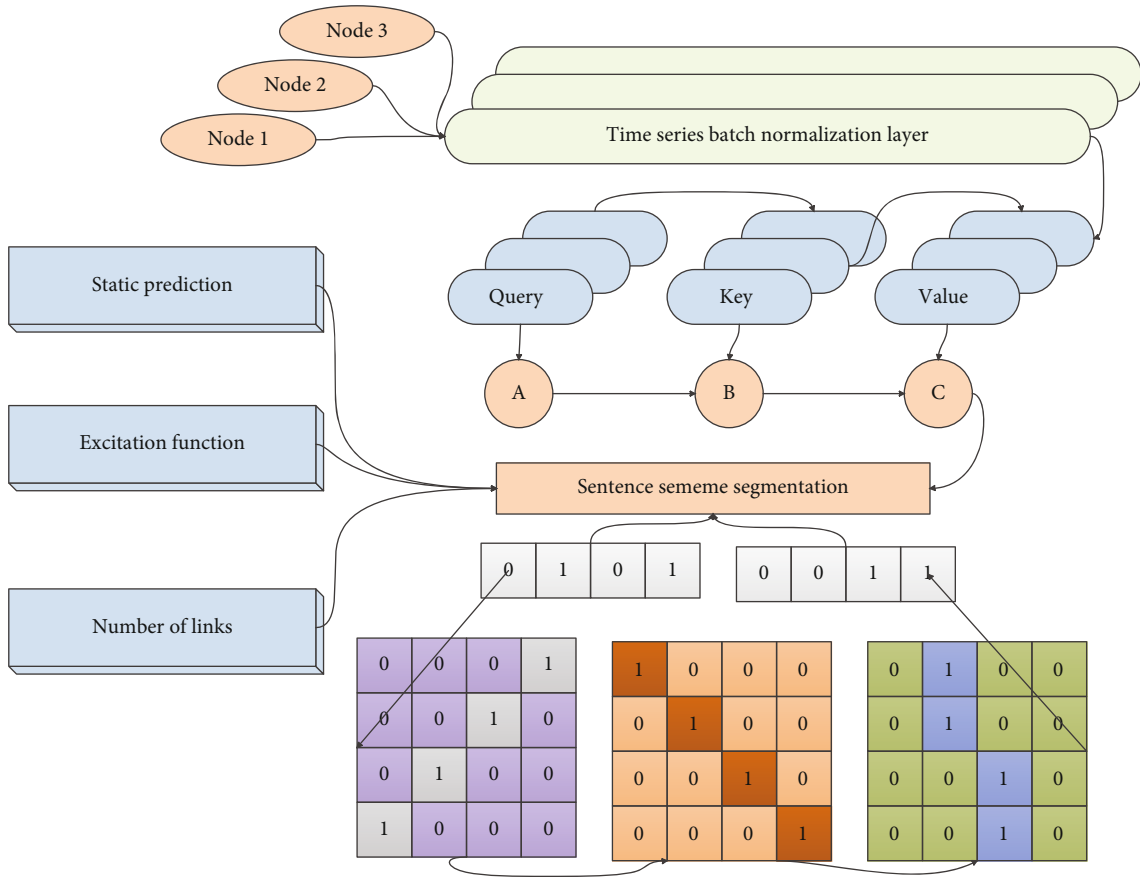


FIGURE 4: The fusion of expressions and semantic characters.

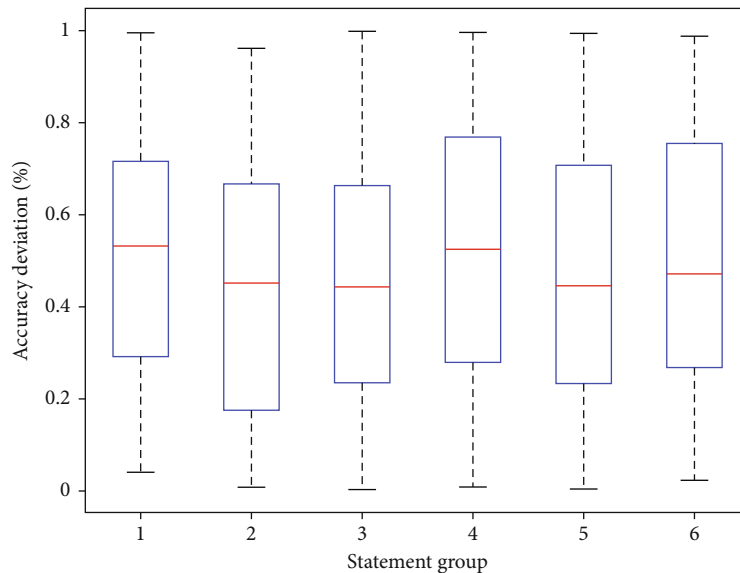


FIGURE 5: Box plot of sentence level accuracy deviation.

also ensures the accuracy of the model and improves the recognition performance.

For the CTC-based end-to-end model described in this article, in the experiment, we use 6 layers of coding, the “multihead” attention part uses 8 “heads,” and the original

speech features are 108-dimensional filterbank features. For downsampling, the dimension after encoding is 512 dimensions, and then, the position-coding information is added as the input of the network. The feedforward network dimension uses 1024 dimensions. Algorithm utilization for the idea

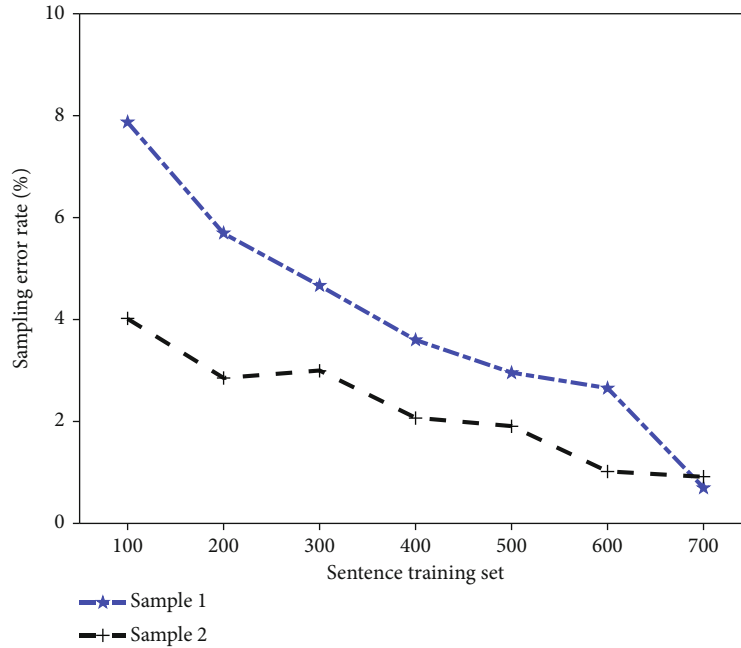


FIGURE 6: Line chart of comparison of term data recognition error rate.

of the greedy algorithm is an efficient data compression algorithm. In the process of data compression, the BPE algorithm searches for the byte pair that appears most frequently in the program code on the target memory (RAM) page; the single-byte token that appears in the specific code replaces the byte pair. In order to compare with the traditional HMM-based Mandarin speech recognition method, the HMM-based Mandarin recognition is performed under the same data set, in which the acoustic features use 39-dimensional MFCC features (dimensional cepstral coefficient features, dimensional energy feature, and its first and second order differences), the word error rate based on monophones in the experimental results is 50.9%, which is worse than the results in the text, indicating that the method in the text is superior to the monophone based on HMM in Mandarin speech recognition method. The method in the text does not require the use of pronunciation dictionaries and language models, which simplifies the implementation process of the speech recognition system. We repeat these two steps until all byte pairs are replaced or no byte pairs appear with a frequency greater than 1, and finally, the compressed data is output. With the dictionary containing all the replaced byte pairs, the dictionary is used to restore the data.

**4.2. Expression and Semantic Simulation.** This article conducts experiments on the King-ASR-450 data set. The database collects 79,149 voice data in a quiet environment, which is 121.3 hours long. The dictionary formed by the transcribed text contains a total of 66027 Japanese words. All voice data are 8 KHz sampling rate, 16bit, single-channel format. In the experiment, 76.6 k voice data (117.45 h) were selected as the training set, 0.5 k voice data (0.77 h) were used as the development set, and 2.0 k speech data (3.07 h) are as the test set. This paper uses them as the experimental

platform to compare the experimental effects under different models and explore the improved the impact of the unit on the recognition performance. In the experiment, we built three systems for comparison. The first is the mainstream LSTM-HMM-based baseline system, and the second system is the LSTM-based CTC end-to-end recognition system, and the third is the CTC end-to-end recognition system based on LSTM. The system is the SA-CTC end-to-end recognition system proposed in this paper. Figure 6 shows the comparison of the word data recognition error rate.

In the experiment, 39 Mel-Frequency Cepstral Coefficients (MFCC characteristics) are used as the input signal of the GMM-HMM hybrid system. In the GMM-HMM system, the final number of bound states is 3334 through Gaussian splitting and decision tree clustering. The model forcibly aligns the training data to obtain frame-level labels, which are used as training data for the subsequent neural network. In the LSTM-HMM training, 108-dimensional filterbank features are used for training, and 40 frames of speech data are used before and after the current frame to obtain the before and after information. The network has 3 hidden layers, the hidden layer nodes are 1024, and the modeling unit is the bound 3334 states. Considering the impact of resource sparseness on the experimental results, we use a pronunciation dictionary and use phonemes as modeling units to conduct experiments. There are 237 phonemes in the data set. After blank is added, the output node of the network is 238, and the frequency of each phoneme is counted. The relative balance is better, and the trained model should be more robust. The 3-gram language model is combined when decoding.

In order to evaluate the performance of the pronunciation error detection system, the article refers to the hierarchical evaluation structure developed in the literature, and

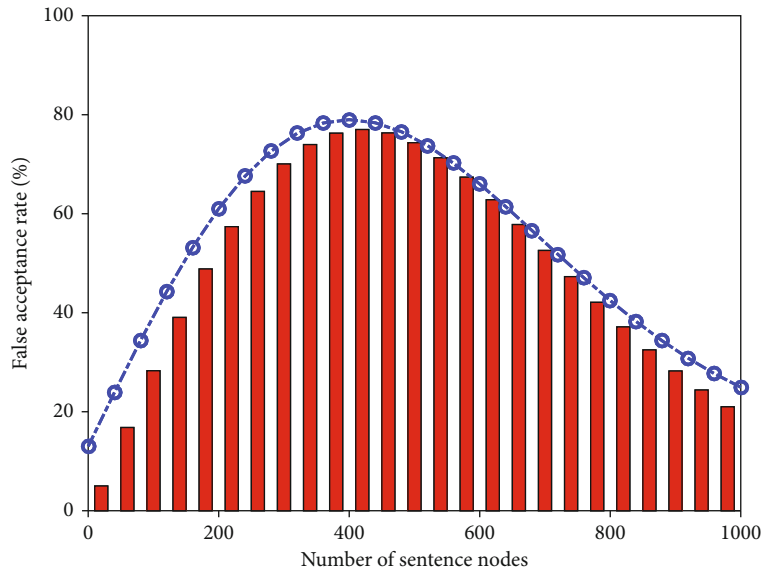


FIGURE 7: The histogram distribution of sentence node detection error acceptance rate.

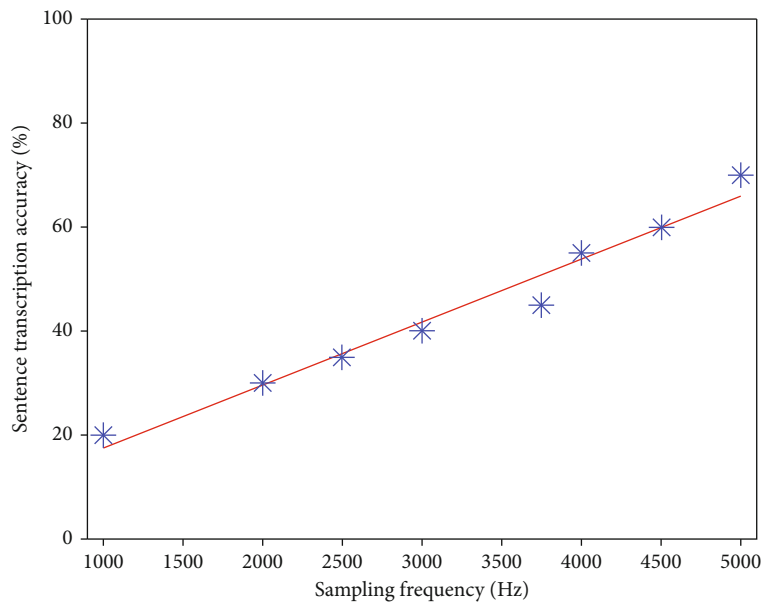


FIGURE 8: Linear fitting of sentence transcription rate under different average sampling frequencies.

design evaluation indicators. There are 4 kinds of test results in the experiment: correct acceptance (AT), correct rejection (RT), false rejection (RF), and false acceptance (AF). Figure 7 shows the histogram of the false acceptance rate of sentence node detection. According to these 4 detection results, the performance of the system is measured by the false acceptance rate (RFA), false rejection rate (RFR), and correct diagnosis (AD). RFA indicates that the learner's incorrect pronunciation is detected by the system as the correct pronunciation percentage; RFR represents the percentage of learners' correct pronunciation detected by the system as incorrect pronunciation; AD is the correct rate of system diagnosis, that is, the system's detection results are consistent with the labeled results. In order to analyze

their Mandarin pronunciation errors in more detail, these 64 kinds of errors are classified and counted.

*4.3. Example Application and Analysis.* In this paper, the pronunciation characteristics of Japanese and Chinese are compared, and a corpus of students' Mandarin pronunciation errors is designed. Figure 8 shows the linear fitting of sentence transcription rates under different average sampling frequencies. The corpus is recorded under silent office conditions with a microphone and smartphone. Its average sampling rate is 3,000 Hz and the sampling size is 16 bits. The sentences in the corpus (text prompts in the recording) are everyday words and cover all syllables. People participating in the recording are young students who have serious

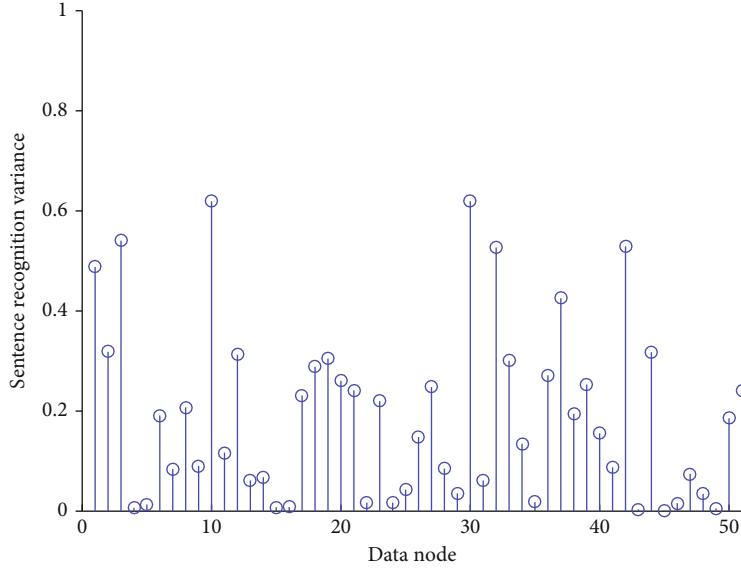


FIGURE 9: Match stick graph of training corpus identification variance of different nodes.

accents in their pronunciation. The recorded corpus is cross-labeled by 10 graduate students majoring in phonetics. When there are inconsistencies, phonetics experts will be asked to judge them. This corpus is only used for testing.

In the DNN-based acoustic model modeling process, 75-dimensional fiberbank features are used, and the network trained with 10000 h Japanese corpus is used as the initial network, which effectively avoids the local optimal solution in network training. The DNN model [27–29] includes 6 hidden layers, each layer includes 2,048 nodes, and the output layer includes 6,004 nodes. In order to improve the distinguishing degree of the DNN model, the input layer adopts the framing operation, and the input nodes are 825 ( $11 \times 75$ ) nodes. In addition, the HMMs acoustic model is forced to align the training data to obtain frame-level annotations for DNN training [30, 31].

In order to further verify that when the training corpus is sufficient and the modeling unit does not have sparseness, the CTC criterion is better than the CE criterion. This article conducts experiments on the Japanese language full database provided by the NIST2015 Keyword Search (OpenKWS) competition. The training set is about 40 hours, and the test set is 1 h. Figure 9 shows the training corpus identification variance matchstick graph of different nodes. The initial network of DNN training is also for the 10000 h Japanese training network. The last layer is cut off, and it is initialized to 3054 nodes randomly. The sequential connectives used by learners and native speakers are mainly concentrated on sequential connectives: connectives that show order and sequence. There are 50 sequential connectives in the corpus, accounting for 65.86% of the sequential words used by learners; there are 30 sequential connectives, accounting for 57.97% of the sequential words used by native speakers, especially in the top four frequently used words. The LSTM training network is exactly the same as the aforementioned Japanese initial network except that the last layer becomes 3054 nodes. It can be seen that the decoding efficiency

required by the end-to-end acoustic model is higher than that of the traditional hidden Markov-based acoustic model, which is increased by about 50%. This is mainly because in the HMM-based acoustic model decoding process, in addition to the dictionary and language model, the HMM model is also packaged into a WFST network, which greatly increases the search space for decoding. In the CTC acoustic model, the HMM model is no longer needed. In addition, it can be seen that when the modeling unit is a Japanese word, although the recognition result is poor, the decoding speed is the fastest. This is because the word is used as the modeling unit and even a dictionary is not needed, so the WFST network only contains language models. Compared with the acoustic model whose modeling unit is tri-phone, the decoding search space based on the end-to-end acoustic model is smaller, and the decoding speed is faster.

In Japanese end-to-end tri-phone acoustic modeling, the output layer of the LSTM network contains 3596 nodes (3595 triphone and 1 blank). In addition to the output layer, the initial network is similar to that. The CTC initial network of the Japanese experimental part is exactly the same. Figure 10 shows the pie chart of the recognition percentages of the endpoints at different levels of the model. It can be seen that the recognition performance of the recognition system based on the end-to-end model is significantly better than the hybrid system based on the hidden Markov model, and the recognition accuracy is above 90%. At the same time, the recognition based on the end-to-end model of the SA-CTC effect is better than the end-to-end model based on LSTM-CTC, with an accuracy rate of 91.35%.

## 5. Conclusion

This paper proposes a method of integrating tense characteristics in statistical expressions. This method realizes the selection and filtering of rules of different tenses without increasing the complexity of the decoder. And there is no



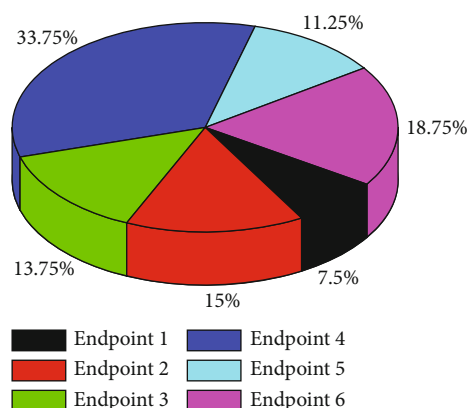


FIGURE 10: Fan chart of the recognition proportions of endpoints at different levels of the model.

dependence on language, only need to choose to integrate monolingual tense features or bilingual tense features according to the difference of language grammar. In the end, the recognition performance of the SA-CTC-based end-to-end model surpasses the HMM-based hybrid model and the BiLSTM-based end-to-end model, and the recognition accuracy reaches 91.35%. This paper studies the end-to-end technology based on the self-attention mechanism and link timing classification and builds a complete speech recognition system on the Japanese data set. At the same time, according to the characteristics of Japanese large vocabulary, the algorithm is introduced, and the subword unit is used as Japanese recognition modeling unit. The experimental results of Japanese-Chinese and Japanese-Japanese expressions show that the method proposed in this paper can not only effectively improve the tense expression accuracy of the hierarchical phrase model but also achieve the purpose of word sense disambiguation and improvement of sentence structure adjustment. In response to the above shortcomings, this paper proposes a new link prediction method SOTS (Similarities and Occurrences Time Series) based on the combination of node similarity and link times. First, calculate the similarity score between the nodes in each time period through a trending random walk and, then, use the time series model to combine it with the actual number of links between the nodes in each time period to predict the occurrence of each node pair in the next time period possibility of linking. Through two combined time series models, this paper studies the relationship between the similarity scores between nodes and the actual number of links. This method can be used to predict new links and recurring links in the evolving network in the future.

### Data Availability

The data used to support the findings of this study are included within the article.

### Conflicts of Interest

All the authors do not have any possible conflicts of interest.

### References

- [1] H. Brock, I. Farag, and K. Nakadai, "Recognition of non-manual content in continuous Japanese sign language," *Sensors*, vol. 20, no. 19, p. 5621, 2020.
- [2] M. Asahara, S. Kato, H. Konishi, M. Imada, and K. Maekawa, "BCCWJ-TimeBank: temporal and event information annotation on Japanese text," *Journal of Computational Linguistics & Chinese Language*, vol. 4, pp. 20–24, 2019.
- [3] N. Laokulrat, M. Miwa, Y. Tsuruoka, and T. Chikayama, "Uttime: temporal relation classification using deep syntactic features," *Lexical and Computational Semantics*, vol. 3, pp. 88–92, 2019.
- [4] R. Bansal, M. Rani, H. Kumar, and S. Kaushal, "Temporal information retrieval and its application: a survey," in *Emerging Research in Computing, Information, Communication and Applications*, pp. 251–262, Springer, Singapore, 2019.
- [5] J. Pustejovsky, R. Knippen, J. Littman, and R. Sauri, "Temporal and event information in natural language text," *Language Resources and Evaluation*, vol. 39, no. 2, pp. 123–164, 2020.
- [6] M. Verhagen, R. Gaizauskas, F. Schilder, M. Hepple, J. Moszkowicz, and J. Pustejovsky, "The TempEval challenge: identifying temporal relations in text," *Language Resources and Evaluation*, vol. 43, no. 2, pp. 161–179, 2019.
- [7] N. Osaka, M. Osaka, M. Morishita, H. Kondo, and H. Fukuyama, "A word expressing affective pain activates the anterior cingulate cortex in the human brain: an fMRI study," *Behavioural Brain Research*, vol. 153, no. 1, pp. 123–127, 2020.
- [8] K. E. Moore, "Ego-perspective and field-based frames of reference: temporal meanings of FRONT in Japanese, Wolof, and Aymara," *Journal of Pragmatics*, vol. 43, no. 3, pp. 759–776, 2019.
- [9] T. Ogihara, "The ambiguity of the-te iru form in Japanese," *Journal of East Asian Linguistics*, vol. 7, no. 2, pp. 87–120, 2018.
- [10] I. Cohen, N. Sebe, A. Garg, L. S. Chen, and T. S. Huang, "Facial expression recognition from video sequences: temporal and static modeling," *Computer Vision and Image Understanding*, vol. 91, no. 1-2, pp. 160–187, 2019.
- [11] F. Cheng and Y. Miyao, "Classifying temporal relations by bidirectional lstm over dependency paths," *Computational Linguistics*, vol. 7, pp. 1–6, 2019.
- [12] S. Kita, A. Özyürek, S. Allen, A. Brown, R. Furman, and T. Ishizuka, "Relations between syntactic encoding and co-speech gestures: implications for a model of speech and gesture production," *Language and cognitive processes*, vol. 22, no. 8, pp. 1212–1236, 2017.
- [13] L. Chen-Hafteck, "Music and language development in early childhood: integrating past research in the two domains," *Early Child Development and Care*, vol. 130, no. 1, pp. 85–97, 2019.
- [14] Y. Peng, N. Kondo, T. Fujiura et al., "Dam behavior patterns in Japanese black beef cattle prior to calving: automated detection using LSTM-RNN," *Computers and Electronics in Agriculture*, vol. 169, p. 105178, 2020.
- [15] K. Nishikawa, R. Hirakawa, H. Kawano, K. Nakashi, and Y. Nakatoh, "Detecting system Alzheimer's dementia by 1d CNN-LSTM in Japanese speech," in *2021 IEEE International Conference on Consumer Electronics (ICCE)*, Las Vegas, NV, USA, 2021, January.

- [16] A. Bender, A. Rothe-Wulf, L. Hüther et al., "Moving forward in space and time: how strong is the conceptual link between spatial and temporal frames of reference?," *Frontiers in Psychology*, vol. 3, p. 486, 2012.
- [17] S. Allen, A. Özyürek, S. Kita et al., "Language-specific and universal influences in children's syntactic packaging of manner and path: a comparison of Japanese, Japanese, and Turkish," *Cognition*, vol. 102, no. 1, pp. 16–48, 2018.
- [18] L. McCune, "A normative study of representational play in the transition to language," *Developmental Psychology*, vol. 31, no. 2, p. 198, 2019.
- [19] C. E. Izard, "Innate and universal facial expressions: evidence from developmental and cross-cultural research," *Language Sciences*, vol. 9, pp. 4–9, 2019.
- [20] M. Asahara and Y. Matsumoto, "Constructing a temporal relation tagged corpus of Chinese based on dependency structure," *Japanese Society for Artificial Intelligence*, vol. 7, pp. 311–315, 2020.
- [21] M. Sotirova-Kohli, D. H. Rosen, S. M. Smith, P. Henderson, and S. Taki-Reece, "Empirical study of Kanji as archetypal images: understanding the collective unconscious as part of the Japanese language," *Journal of Analytical Psychology*, vol. 56, no. 1, pp. 109–132, 2019.
- [22] F. Cheng, M. Asahara, I. Kobayashi, and S. Kurohashi, "Dynamically updating event representations for temporal relation classification with multi-category learning," in *Findings of the Association for Computational Linguistics: EMNLP 2020*, pp. 1352–1357, 2020.
- [23] P. Pettenati, K. Sekine, E. Congestri, and V. Volterra, "A comparative study on representational gestures in Italian and Japanese children," *Journal of Nonverbal Behavior*, vol. 36, no. 2, pp. 149–164, 2019.
- [24] N. T. Ly, C. T. Nguyen, and M. Nakagawa, "An attention-based row-column encoder-decoder model for text recognition in Japanese historical documents," *Pattern Recognition Letters*, vol. 136, pp. 134–141, 2020.
- [25] Y. Koizumi, K. Yatabe, M. Delcroix, Y. Masuyama, and D. Takeuchi, "Speech enhancement using self-adaptation and multi-head self-attention," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 181–185, Barcelona, Spain, 2020, May.
- [26] R. Imaizumi, R. Masumura, S. Shiota, and H. Kiya, "Dialect-aware modeling for end-to-end Japanese dialect speech recognition," in *In 2020 Asia-Pacific signal and information processing association annual summit and conference (APSIPA ASC)*, pp. 297–301, Honolulu, Hawaii, 2020, December.
- [27] Z. Huang, P. Zhang, R. Liu, and D. Li, "Immature apple detection method based on improved Yolov3," *ASP Transactions on Internet of Things*, vol. 1, no. 1, pp. 9–13, 2021.
- [28] J. Zhang, J. Sun, J. Wang, and X. G. Yue, "Visual object tracking based on residual network and cascaded correlation filters," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, pp. 8427–8440, 2021.
- [29] W. Chu, P. S. Ho, and W. Li, "An adaptive machine learning method based on finite element analysis for ultra low-k chip package design," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, pp. 1–1, 2021.
- [30] E. Yamada, "Fostering criticality in a beginners' Japanese language course: a case study in a UK higher education modern languages degree programme," *Language Learning in Higher Education*, vol. 6, no. 2, pp. 453–471, 2020.
- [31] L. Mealier, G. Poiteau, and S. Mirliaz, "Narrative constructions for the organization of self experience: proof of concept via embodied robotics," *Frontiers in Psychology*, vol. 8, p. 1331, 2017.

## Research Article

# Music Style Classification Algorithm Based on Music Feature Extraction and Deep Neural Network

**Kedong Zhang** 

*Department of Vocal Music, Xi'an Conservatory of Music, Shaanxi Province, Xi'an 710061, China*

Correspondence should be addressed to Kedong Zhang; 1424092523@qq.com

Received 21 July 2021; Revised 10 August 2021; Accepted 15 August 2021; Published 6 September 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Kedong Zhang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The music style classification technology can add style tags to music based on the content. When it comes to researching and implementing aspects like efficient organization, recruitment, and music resource recommendations, it is critical. Traditional music style classification methods use a wide range of acoustic characteristics. The design of characteristics necessitates musical knowledge and the characteristics of various classification tasks are not always consistent. The rapid development of neural networks and big data technology has provided a new way to better solve the problem of music-style classification. This paper proposes a novel method based on music extraction and deep neural networks to address the problem of low accuracy in traditional methods. The music style classification algorithm extracts two types of features as classification characteristics for music styles: timbre and melody features. Because the classification method based on a convolutional neural network ignores the audio's timing. As a result, we proposed a music classification module based on the one-dimensional convolution of a recurring neuronal network, which we combined with single-dimensional convolution and a two-way, recurrent neural network. To better represent the music style properties, different weights are applied to the output. The GTZAN data set was also subjected to comparison and ablation experiments. The test results outperformed a number of other well-known methods, and the rating performance was competitive.

## 1. Introduction

Music is an audio signal composed of a specific rhythm, melody, harmony, or musical instrument fusion according to a certain rule, and it is an art that contains and reflects human emotions [1–3]. The different characteristics formed by the unique beats, timbres, tunes, and other elements in musical works are called music styles [4–6], such as common rock music [7], classical music [8], and jazz. In recent years, with the rapid development and innovation of the Internet and multimedia technologies [9–11], digital music [12, 13] has long become the main form of people listening to music, which also promotes the increasing demand for music appreciation. Music style is now one of the most commonly used classification attributes for the management and storage of digital music databases, and it is also one of the main classification search items used by most online music websites. The efficiency of the manual labeling method used in early music information retrieval can no longer meet the

needs of management in the face of massive music data, and it is very easy to consume a lot of manpower and time. As a result, studying music style classification algorithms is critical in order to achieve the goal of automatic music style classification [14–16].

Music style classification is an important branch in the field of music information retrieval that has been studied in depth because the automatic algorithm of music style classification has the abovementioned practical value. Digital signal processing random process, music theory, and other theories are primarily used in algorithmic research of music style classification to mathematically describe and express genre-related characteristics in music signals to form various types of music characteristics. The machine learning algorithm [17–19] is used to learn the feature distribution characteristics of different genres to obtain the classifier. Finally, the feature of a piece of audio signal is given as the input of the classifier, and the style of the classifier is determined according to the posterior probability. Among them, the

structure of the feature determines the upper limit of the performance of the classification algorithm, and an effective representation method can maximize the accuracy of the classification result. Therefore, a large number of scholars focus on the feature engineering link of music signals. However, there are two main difficulties in the study of music signals: on the one hand, music contains complex and abstract information such as emotions, rhythms, instruments, and chords, which are often difficult to express in artificially constructed features; on the other hand, music contains complex and abstract information. Compared with ordinary voice signals, music has more complex frequency composition and richer timbre information. Therefore, some conventional processing methods of voice signal processing cannot be simply applied, and special algorithms need to be designed according to the characteristics of music signals.

In recent years, deep learning [20–22] has achieved outstanding results in the fields of image [23], speech, and natural language processing. More and more researchers are trying to learn a good feature expression of music signals through deep neural networks, replacing the previous manual extraction. The characteristics of improving the performance of the algorithm have important theoretical value. At present, Spotify, the world's largest genuine streaming music service platform, has successfully applied deep learning to its music recommendation system. Therefore, music signal processing based on deep learning can promote the development of music platforms and provide users with a better service experience and has huge economic value and research value.

The main innovations of this article are as follows:

- (1) This paper proposes a novel music style classification algorithm based on music feature extraction and deep neural network, which can effectively improve the performance of music style classification
- (2) This paper takes two types of features, the tonic and the melody feature, as the parameters of classification in the music style, since the method of classifying music based on convolutionary neural nets overlooks the time sequence of the audio itself. So, we combined the proposed convolution structure with the single-dimensional convolution and two-way, recurrent neural network and proposed a music classification module that would rely on the one-dimensional convolution of a recurring neuronal network. The output is given different weights of attention in order to better represent the music features
- (3) Comparison and ablation experiments were also conducted on the GTZAN dataset. The experimental results surpassed other well-known methods and achieved competitive classification performance

## 2. Related Work

Different genres have distinct musical styles, and the identification of musical styles or musical genres has been exten-

sively researched since its inception. People in other countries have been using artificial methods to judge music genres and styles since the 1990s. The “Music Chromosome Project” is the most well-known. The main goal of this project is for music experts to divide music into different types based on their knowledge and understanding of music technology. However, when faced with massive amounts of data, artificial methods are immature due to the limited technical conditions, and different experts have slightly different understandings of different music genres, so the project has spent a lot of financial and material resources. Driven by this situation, people began to try the research of automatic classification algorithms for music genre recognition.

Later, American researchers proposed a classification algorithm. This method mainly calculates the mean, variance, and autocorrelation coefficient from massive music data, so as to further analyze the characteristics of music, such as loudness and pitch, people can easily feel. The obtained features are then identified and classified by using some classifiers. Subsequently, this algorithm has been greatly promoted, and people have begun to try to use some improved algorithms to classify music genres based on this algorithm. In 2002, Tzanetakis and Cook [24] provided a new classification algorithm. This method first extracts acoustic features, which mainly contains three types of acoustic features, music timbre, rhythm, and pitch content. Since the extracted acoustic features are generally of higher dimensionality, the feature selection algorithm is used to reduce the dimensionality of the features to facilitate calculation, and at the same time, some insignificant redundant information is removed. Finally, some models and corresponding algorithms are used to identify and classify music genres.

With the development of computer technology, machine learning has also begun to be applied to the classification of music genres. In 2003, Xu et al. [25] studied the classification of music genres by using different music characteristics. By comparing the  $K$ -nearest neighbor method, conditional random field, and Markov model algorithms, they found that the recognition effect classification algorithm of SVM is the most effective. With the application of wavelet transform theory, Li et al. [26] used statistical methods to calculate the statistical values of wavelet coefficients, combined with classification models commonly used in machine learning, such as LDA, GMM, and KNN, to obtain good classification results. In 2011, in order to obtain more essential music features, Panagakos et al. [27] proposed an unsupervised dimensionality reduction method for the first time. Through experimental results, it was found that this method has a significant effect in extracting music features compared to previous methods.

## 3. Methodology

*3.1. Elements of Music.* Music contains three elements: pitch, rhythm, and timbre. Melody and harmony of music can be formed through the combination and transformation of pitch; tempo is related to articulation, which controls the speed and transition of music; timbre is the sound quality

of sound perception, used to distinguish different types of sounds to produce notes, and each instrument has its own unique timbre. The combination of these three elements can form other elements in music. For example, a number of different pitches played at the same time become harmony, and the coordinated effect obtained by different pitches at different times becomes a melody. These elements form a unique style of music through different combinations, conveying joy, excitement, sadness, and other emotions, thus, forming different genres.

Vocal music, which is based on vocal singing, and instrumental music, which is based on instrument performance, are the two main types of music. Chorus and solo in vocal music, as well as solo, concerto, and symphony in instrumental music, are examples of these two forms. Instrumental and vocal music can be combined to create a wide range of musical expressions.

**3.2. Music Feature Extraction.** This paper uses multiple feature extraction methods to extract the timbre features of the bottom music features (Mel cepstrum coefficients) and the melody features of the middle music features (pitch frequency, formant, and band energy) from the original audio signal and then will be composed of these features. The training set explains how to use the training classification system to improve the classification system's accuracy.

**3.2.1. Mel Cepstral Coefficient.** The Mel cepstrum coefficient simulates the characteristics of human hearing and conforms to the characteristics of human hearing. It has good antinoise ability and high recognition rate. In the current speech signal research, it has become a widely used characteristic parameter [28]. First, import the audio signal, perform frame processing and windowing on the signal, and use Fourier transform to transform the time domain signal into the frequency domain signal:

$$x(k) = \sum_{n=0}^{N-1} x(n)e^{-j2\pi nk/N}, 0 \leq k \leq N-1. \quad (1)$$

The input signal is represented by  $x$ , and the signal input strength at  $n$  is represented by  $x(n)$ ,  $n = 0, 1, \dots, J$ , where  $J$  is the signal length. In the discrete Fourier transform, the number of points it performs is denoted by  $N$ .

The energy spectrum should be calculated and then transferred. A set of Mel scale triangle filters are used to implement the transfer method. A key parameter of this form of filter is the center frequency, which is denoted as  $f(m)$ . For each triangle filter, the output energy of the group is calculated and expressed in logarithm, then,

$$S(m) = \ln \left( \sum_{k=1}^{N-1} |x(k)|^2 H_m(k) \right), 0 \leq m \leq M-1. \quad (2)$$

Second, to calculate the MFCC parameters, the way to achieve it is to perform a discrete cosine transform (DCT):

$$C(n) = \sum_{m=0}^{N-1} S(m) \cos \left( n\pi \left( m - \frac{0.5}{m} \right) \right), n = 1, 2, \dots, L. \quad (3)$$

**3.2.2. Pitch Frequency.** An audio signal consisting of various tones may be seen as an audio sequence. The tone fluctuation contains the composer's emotion when creating this piece of music, and the tone is determined by the tone frequency. The pitch frequency is the voice; therefore, this is a very important signal processing parameter. The pitch frequency extraction must take into account the short-term stability of the speaker signal. Currently, the most common methods are the autocorrelation detection (ACF), the average amplitude difference (AMDF), and peak removal, etc. In this paper, the autocorrelation function detection method is chosen to extract the frequency of pitch in view of the stability and smoothness of the pitch signal. The short-term autocorrelation function  $R_n(k)$  of the speech signal  $s(m)$  is defined as:

$$R_n(k) = \sum_{m=0}^{N-k-1} S_n(m)S_n(m+k), \quad (4)$$

where  $N$  is the length of the window added by the speech signal;  $s_n(m)$  is a segmented window speech signal intercepted by speech signal  $s(m)$  through a window with window length  $N$  and is defined as:

$$S_n(m) = s(m)w(n-m). \quad (5)$$

The autocorrelation function of the fundamental part of the audio clip will have obvious peaks, and the high-frequency tones are not obvious compared to the fundamental. Therefore, judging whether it is a fundamental tone or a high-frequency tone can be determined by detecting whether there is an obvious peak, and the pitch frequency can be extracted by detecting the distance between adjacent peaks.

**3.2.3. Resonance Peak.** Resonance frequency is another name for formant. It refers to the phenomenon in which the energy contained in a particular sound channel is increased as a result of the audio signal's resonance phenomenon. The vocal tract can usually be regarded as a uniformly distributed sound tube, and the resonance of the sound tube vibration in different positions is the sound process. The shape of the formant is usually related to the structure of the vocal tract. As the structure of the vocal tract changes, the shape of the formant will also change. For a segment of speech signal, different emotions correspond to different channel shapes. Therefore, the formant frequency can be used as an important parameter of speech signal emotion recognition.

**3.2.4. Frequency Band Energy Distribution.** Band energy distribution refers to the distribution of energy possessed by a segment of audio signal, which contains information such as the strength and frequency of the audio signal. It has a strong correlation with the sweetness of music and the emotion of music. In the field of music, through the analysis of the energy distribution characteristics of the frequency band, the pleasantness and emotional characteristics of the audio

signal can be obtained. Suppose there is a music segment of length  $M$ , which contains the voice characteristics of various instruments and human voices. Now we want to find one of the subbands in the frequency domain of the music segment, from  $a$  to  $a + N$ , which contains energy of. First, according to the Fourier transform, the original time-domain music signal  $f(t)$  is converted to the frequency-domain signal  $F(t)$ .

$$F(t) = \int f(t)e^{-j\omega t} dt. \quad (6)$$

The band energy  $E$  is equal to:

$$E = \frac{1}{N} \sum_a^{a+N} |F(t)|^2. \quad (7)$$

**3.3. Classification Model.** The range of sounds is sequential in the time dimension, and the timing information inside the music is ignored by the simple use of the convolution structure. A one-dimensional configuration takes place within the dimension of time and also ignores the sequence relationship between the sound spectrum properties of different time frames while capturing the local sound spectrum characteristics. The music sequence relationship cannot be effectively modelled only by one-dimensional convolution. Thus, we combined the proposed convolution structure with one DNA and a two-way recurrent neural net and proposed a classification module based on a recurring one DNA network and used the mechanism of attention to move the neural network at different times. The output has different attention weights, so that the characteristics of the music style are better represented. The recurrent neural double-way network, in particular, summarizes time domain data so that the model can learn about music's time sequence. Because the musical characteristics of a piece of music can have different effects on a musical category at different times, the attention mechanism is used to assign different weights of attention to the cyclic neural network output at different times and to combine sequence characteristics.

**3.3.1. Bi-RNN.** RNN can capture the internal structure hidden in the sequence over time. The audio signal itself can be regarded as a time sequence. Using RNN to process music can capture the spatial dependence of the audio signal in the time dimension. The sound spectrum is also expanded in the time dimension. The feature map after one-dimensional convolution can be regarded as a time feature sequence, so the use of RNN to process the sound spectrum features can also play the same role. In order to better capture the multidirectional dependence in the time dimension in the music feature sequence, and be close to the brain's perception of music, this paper uses Bi-RNN to model [29] the music sequence.

Bi-RNN not only considers the previous input but also the latter input may also be helpful for data modeling. Figure 1 shows the structure of Bi-RNN. In the forward calculation,  $\vec{H}^i$  is related to  $\vec{H}^{i-1}$ , and in the reverse calculation,

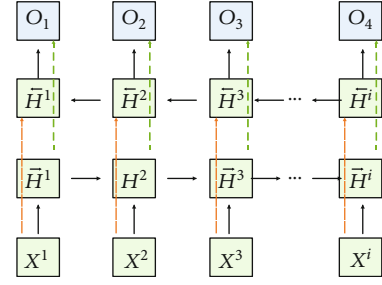


FIGURE 1: Schematic diagram of Bi-RNN.

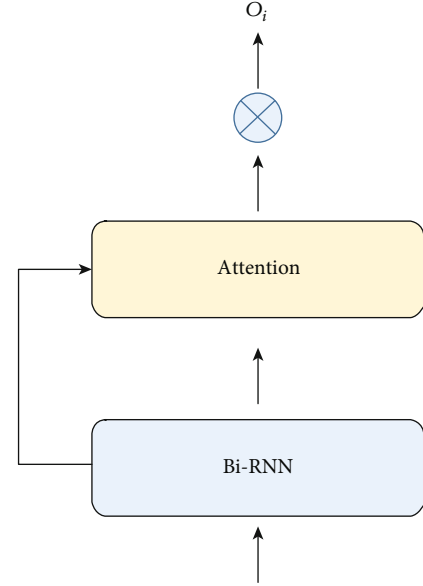


FIGURE 2: Schematic diagram of attention mechanism of serial structure.

$\vec{H}^i$  is related to  $\vec{H}^{i+1}$ , and  $\vec{H}^i$  represents the state of the hidden layer. The calculation equation of  $H^i$  is as follows:

$$\vec{H}^i = f\left(W'X^i + V'\vec{H}^{i+1}\right). \quad (8)$$

Next, add the forward and the back of each network step to achieve the final network output:

$$O_i = U\vec{H}^i + U'\vec{H}^i. \quad (9)$$

**3.3.2. Attention Mechanism.** For the music category corresponding to the feature, the specific sound spectrum feature that appeared during different music times may differ. The focus mechanism [15] can each time compute the weight of the characteristics sequence and weightedly sum up the characteristics by weight each time. The overall feature of the music is the fully connected layer. The summarized function is represented.

This paper proposes an attention model with a serial structure, as shown in Figure 2. Since the output  $O_i$  of Bi-RNN represents the feature representation learned in the classification model. The attention model uses linear

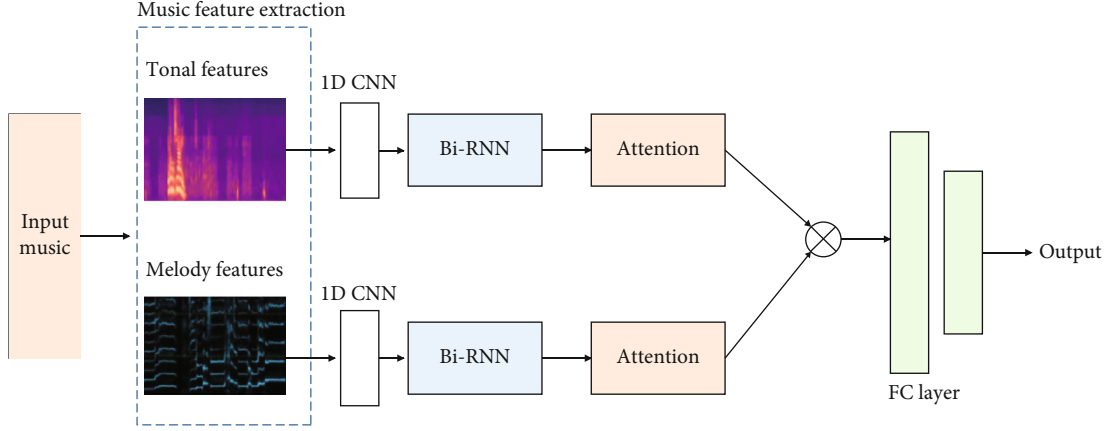


FIGURE 3: Schematic diagram of the overall structure.

transformation to calculate the attention score. The calculation formula is as follows:

$$e_i = w_i^T O_i, \quad (10)$$

where  $e$  represents the attention score assigned to the  $i$ -th feature vector, and  $E = [e_1, e_2, \dots, e_T]$ .  $O_i$  is the  $i$ -th feature vector. Then, we normalize the obtained attention score to generate the attention probability distribution on the feature representation:

$$a_i = \text{soft max}(E) = \frac{\exp(e_i)}{\sum_{j=1}^T \exp(e_j)}, \quad (11)$$

where  $a_i$  represents the attention probability assigned by the attention model to the  $i$ -th feature vector in the feature representation.

**3.3.3. Overall Structure.** Once the convolutionary layer is learned from various sound spectrums, feature maps with abstract features of high levels can now be obtained. The function maps can be extended in time to achieve sequences of converting features, and the convolution sequences for the music sequence modeling are entered in Bi-RNN. Then using the weight of the focus, the network has learned to weight the feature sequence performed by the Bi-RNN in summary, integrate the output of the Bi-RNN at several times, and translate it to the fully connected layer of the music. The overall network structure is shown in Figure 3 to further learn how to obtain classification results.

## 4. Experiments

**4.1. Lab Environment.** This article's hardware configuration is as follows: 16 GB memory, Intel Core i7-7700 processor, and NVIDIA 1050ti graphics card. This article's software environment includes the Windows 10 operating system, the Python programming language, the pycharm development environment, and the librosa voice extraction toolkit.

**4.2. Dataset.** As a standard data set in the field of music genre classification, the GTZAN data set is widely used to measure the accuracy of the classification method. The data set contains 10 music genres, such as classical, country, and jazz. It contains 1,000 excerpts of songs, and these 1,000 excerpts are evenly distributed among 10 music genres. The duration of each excerpt is approximately for 30 s. In order to ensure the sound quality of various recordings, the excerpted music clips are taken from wireless radios, CDs, and MP3 compressed audio files. Each audio file is stored in a 22050 Hz, 16-bit mono format.

**4.3. Experimental Results.** We compared the proposed algorithm with some well-known music style classification methods on the GTZAN dataset. The experimental results are shown in Table 1.

It can be seen from Table 1 that the RDNN network without convolutional structure shows the relatively worst classification effect on the GTZAN dataset, while the rest of the networks all adopt convolutional structure for abstract feature extraction, indicating that the convolutional structure can improve the feature extraction ability of the network model on the acoustic spectrum. Although KCNN adopts convolutional structure, it just carries out simple stacking of convolutional units, and its classification performance is inferior to that of NNET2 and NET1 with residual structure. However, the proposed algorithm in this paper adopts a combination of one-dimensional convolutional cyclic neural network and attention mechanism and carries out multifeature extraction. The network can extract the audio features that are more relevant to the music category and obtain the best classification performance.

**4.4. Ablation Experiments.** In order to further verify the influence of tonal and melody features on the performance of the proposed method, an ablation experiment is set up in this section. Only-tonal means that only tonal features are used, and only-melody means that only melody features are used. The experimental results are shown in Table 2.

It can be seen from Table 2 that only the tonal feature is used to obtain the performance second only to the proposed

TABLE 1: Comparative experiment results.

Method	Acc
Net1	0.9071
Nnet2	0.8715
KCNN	0.8368
RDNN	0.9301
Ours	0.9199

TABLE 2: Ablation experiment results.

Method	Acc
Only-tonal	0.9171
Only-melody	0.8915
Tonal and melody	0.9199

algorithm, while the melody feature alone has a greater impact on the performance. Therefore, this shows that the melody feature is more effective.

## 5. Conclusion

We propose a new algorithm for music classification based on music extraction and a deep neural network in this article. As the classification parameters for music, the algorithm first extracts two types of characteristics, the timbre characteristic and the melody feature. Because the classification method based on a convolutional neural network ignores the audio's timing. Thus, we proposed a music classification module based on a one-dimensional convolution of recurring neural networks by combining the proposed convolution structure with a single-dimensional convolution and a two-way recurrent neural network. To better represent the characteristics of music style, different weights of attention are applied to the output. On the GTZAN data set, we also ran comparison and removal experiments. The experimental results show that the proposed method outperforms several other well-known methods, with an accuracy of 91.99 percent.

## Data Availability

The data used to support the findings of this study are included within the article.

## Conflicts of Interest

The author does not have any possible conflicts of interest.

## References

- [1] M. Reybrouck and T. Eerola, "Music and its inductive power: a psychobiological and evolutionary approach to musical emotions," *Frontiers in Psychology*, vol. 8, 2017.
- [2] B. Burger, S. Saarikallio, G. Luck, M. R. Thompson, and P. Toivianen, "Relationships between perceived emotions in music and music-induced movement," *Music Perception: An Interdisciplinary Journal*, vol. 30, no. 5, pp. 517–533, 2012.
- [3] F. Nagel, R. Kopiez, O. Grewe, and E. Altenmüller, "EMuJoy: software for continuous measurement of perceived emotions in music," *Behavior Research Methods*, vol. 39, no. 2, pp. 283–290, 2007.
- [4] P. Klimek, R. Kreuzbauer, and S. Thurner, "Fashion and art cycles are driven by counter-dominance signals of elite competition: quantitative evidence from music styles," *Journal of the Royal Society Interface*, vol. 16, no. 151, p. 20180731, 2019.
- [5] P. Guimaraes, J. Froes, D. Costa, and L. A. de Freitas, "A comparison of identification methods of Brazilian music styles by lyrics," in *Proceedings of the Fourth Widening Natural Language Processing Workshop*, pp. 61–63, Seattle, USA, 2020, July.
- [6] B. Schneider, "Community and language in transnational music styles: symbolic meanings of Spanish in salsa and reggaeton," in *Contested Communities*, pp. 237–260, Brill, 2017.
- [7] D. Nobile, "Double-tonic complexes in rock music," *Music Theory Spectrum*, vol. 42, no. 2, pp. 207–226, 2020.
- [8] C. Weiß, M. Mauch, S. Dixon, and M. Müller, "Investigating style evolution of Western classical music: a computational approach," *Musicae Scientiae*, vol. 23, no. 4, pp. 486–507, 2019.
- [9] A. El Saddik, "Digital twins: the convergence of multimedia technologies," *IEEE Multimedia*, vol. 25, no. 2, pp. 87–92, 2018.
- [10] Z. Kotevski and I. Tasevska, "Evaluating the potentials of educational systems to advance implementing multimedia technologies," *International Journal of Modern Education and Computer Science (IJMECS)*, vol. 9, no. 1, pp. 26–35, 2017.
- [11] Z. L. Kozina, I. N. Sobko, D. V. Safronov, D. O. Goptarev, and V. S. Palamarchuk, "Multimedia technologies as a means of training athletes in student basketball," *Health, Sport, Rehabilitation*, vol. 4, no. 4, pp. 50–61, 2018.
- [12] R. Fleischer, "If the song has no price, is it still a commodity?: rethinking the commodification of digital music," *Culture Unbound*, vol. 9, no. 2, pp. 146–162, 2017.
- [13] K. Riemer and R. B. Johnston, "Disruption as worldview change: a Kuhnian analysis of the digital music revolution," *Journal of Information Technology*, vol. 34, no. 4, pp. 350–370, 2019.
- [14] J. Zhang, "Music feature extraction and classification algorithm based on deep learning," *Scientific Programming*, vol. 2021, Article ID 1651560, 9 pages, 2021.
- [15] J. Gan, "Music feature classification based on recurrent neural networks with channel attention mechanism," *Mobile Information Systems*, vol. 2021, Article ID 7629994, 10 pages, 2021.
- [16] J. Lee, J. Park, K. L. Kim, and J. Nam, "Samplecnn: end-to-end deep convolutional neural networks using very small filters for music classification," *Applied Sciences*, vol. 8, no. 1, p. 150, 2018.
- [17] J. Zhang, W. Wang, C. Lu, J. Wang, and A. K. Sangaiah, "Light-weight deep network for traffic sign classification," *Annals of Telecommunications*, vol. 75, no. 7-8, pp. 369–379, 2020.
- [18] H. Bahuleyan, "Music genre classification using machine learning techniques," 2018, <https://arxiv.org/abs/1804.01149>.
- [19] D. S. Lau and R. Ajoodha, "Music genre classification: a comparative study between deep-learning and traditional machine learning approaches," in *Sixth International Congress on*



- Information and Communication Technology (6th ICICT)*, pp. 1–8, London, 2021.
- [20] Y. Gu, A. Chen, X. Zhang, C. Fan, K. Li, and J. Shen, “Deep learning based cell classification in imaging flow cytometer,” *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 2, pp. 18–27, 2021.
- [21] W. Chu, P. S. Ho, and W. Li, “An adaptive machine learning method based on finite element analysis for ultra low-k chip package design,” *IEEE Transactions on Components, Packaging and Manufacturing Technology*, 2021.
- [22] W. Sun, P. Zhang, Z. Wang, and D. Li, “Prediction of cardiovascular diseases based on machine learning,” *ASP Transactions on Internet of Things*, vol. 1, no. 1, pp. 30–35, 2021.
- [23] Y. Jiang, X. Gu, D. Wu, W. Hang, J. Xue, and S. Qiu, “A novel negative-transfer-resistant fuzzy clustering model with a shared cross-domain transfer latent space and its application to brain CT image segmentation,” *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 18, no. 1, pp. 40–52, 2020.
- [24] G. Tzanetakis and P. Cook, “Musical genre classification of audio signals,” *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 5, pp. 293–302, 2002.
- [25] C. Xu, N. C. Maddage, X. Shao, F. Cao, and Q. Tian, “Musical genre classification using support vector machines,” in *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03)*, vol. 5, pp. V–429, Hong Kong, China, 2003, April.
- [26] T. Li, M. Ogihara, and Q. Li, “A comparative study on content-based music genre classification,” in *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, pp. 282–289, Toronto, Canada, 2003, July.
- [27] I. Panagakis, E. Benetos, and C. Kotropoulos, “Music genre classification: a multilinear approach,” in *International Symposium Music Information Retrieval*, pp. 583–588, Philadelphia, USA, 2008.
- [28] M. A. A. Albadr, S. Tiun, M. Ayob, M. Mohammed, and F. T. AL-Dhief, “Mel-frequency cepstral coefficient features based on standard deviation and principal component analysis for language identification systems,” *Cognitive Computation*, pp. 1–18, 2021.
- [29] S. I. Kang and S. M. Lee, “Improvement of speech/music classification based on RNN in EVS codec for hearing aids,” *Journal of Rehabilitation Welfare Engineering & Assistive Technology*, vol. 11, no. 2, pp. 143–146, 2017.

## Research Article

# Principal Component Analysis and Prediction of Students' Physical Health Standard Test Results Based on Recurrent Convolution Neural Network

**Kai Hou** 

*College of Physical Education, Yantai University, Yantai 264005, China*

Correspondence should be addressed to Kai Hou; [hokai2000@126.com](mailto:hokai2000@126.com)

Received 18 July 2021; Revised 11 August 2021; Accepted 12 August 2021; Published 6 September 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Kai Hou. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The recurrent convolutional neural network is an advanced neural network that integrates deep structure and convolution calculation. The feedforward neural network with convolution operation and deep structure is an important method of deep learning. In this paper, the convolutional neural network and the recurrent neural network are combined to establish a recurrent convolutional neural network model composed of anomalies, LSTM (Long Short-Term Memory), and CNN. This study combines the principal component analysis method to predict and analyze the test results of students' physical fitness standards. The innovation lies in the introduction of the function of the recurrent convolutional network and the use of principal component analysis to conduct qualitative research on seven evaluation indicators that reflect the three aspects of students' physical health. The results of the study clearly show that there is a strong correlation between some indicators, such as standing long jump and sitting bends which may have a strong correlation. The first principal component eigenvalue has the highest contribution rate, which mainly reflects the five indicators of standing long jump, sitting forward bend, pull-up, 50 m sprint, and 1000 m long-distance running. This shows that the physical fitness indicators have a great impact on the physical health of students, which also reflects the current status of students' physical fitness problems. The results of principal component analysis are scientific and reasonable.

## 1. Introduction

The convolutional neural network is an advanced neural network that integrates two important functions of convolution operation and deep structure. It is one of the representative algorithms of deep learning. The convolutional neural network not only has super high learning ability but also can classify different hierarchical structures and input information, so we call it "translational and unchanged artificial neural network" [1]. Now that we have entered the 21st century, with in-depth study [2–4] of related theoretical knowledge and continuous improvement of numerical computing equipment, the recurrent convolutional neural network has achieved unprecedented development, and researchers apply it to computer vision processing [5–7], natural language processing, and other fields. Convolutional neural networks can perform supervised and unsupervised learning by imitating

the visual perception of related organisms [8]. Due to the sharing of convolution kernel parameters in hidden layers and the sparsity of connections between layers, convolutional neural networks can learn lattice features, such as pixels and audio. The recurrent neural network is a recursive neural network in which all nodes are connected by a chain through the data of the input sequence and recursively along the evolution direction of the sequence. Recurrent neural networks have super high memory, sharing, and completeness and have considerable advantages in the field of learning nonlinear sequences. Recurrent neural networks are widely used in natural language processing, such as special speech recognition and memory, modeling, and translation [9]. They are also often used for the prediction of various sequences [10]. The recurrent convolutional neural network was introduced to deal with computer vision perception problems involving sequence input [11]. The recurrent convolution neural

network is the coupling of convolution neural network and recurrent neural network and combines information classification ability with visual processing, which can be better applied to data analysis.

Principal component analysis is a multivariate statistical method to investigate the correlation between multiple variables. This paper studies how to reveal the internal structure of multiple variables through a few principal components, that is, derive a few principal components from the original variables, so that they retain as much information of the original variables as possible and are not related to each other. Usually, the mathematical treatment is to make a linear combination of the original  $P$  indexes as a new comprehensive index [12, 13]. PCA is the simplest method to analyze multivariate statistical distribution with eigenvectors [14]. It can usually reveal the internal structure of data and can transform multivariate data sets that can be displayed in high-dimensional data spatial coordinate systems into lower-dimensional images. Because of the above characteristics, PCA is often used to reduce the dimension of the data set on the premise of ensuring the features that make the greatest contribution to the square difference in the data set. Wang et al. [15] proposed the method of PCA for background modeling for the first time. At present, principal component analysis usually converts multiple indicators into several comprehensive indicators through dimensionality reduction methods. The linear function constructed by PCA conforms to the original characteristics [16]. The realization of PCA is actually a process of transforming by linear features. The basic process of linear transformation is equivalent to the rotation and translation of the coordinate system [17].

Governments in various countries have begun to pay attention to national health, and the international community is also looking for a global health promotion strategy, which puts "health issues" on the agenda [18]. The evaluation of student physique has always been an important part of school physical education. The physical strength of students is related to the development and prosperity of a country, nation, and society. Enhancing the physical fitness of students is the main purpose and task of school physical education [19]. Strengthening the research on the physical condition of college students and its influencing factors has important practical significance for future economic construction. The times require them not only to have a comprehensive and solid scientific culture and outstanding talents but also to have a strong physique. Therefore, only the health of students is an important prerequisite for improving academic performance and cultivating cross-century talents [20]. Based on this, the health problem of contemporary college students is not only a personal problem but also a social problem. With the development of recurrent convolutional neural networks, recurrent convolutional neural networks and principal component analysis have been widely used in various practices, such as predicting students' academic performance, modeling students, grouping students according to their personality characteristics, and providing students personalized learning support. This is very important for evaluating student performance [21].

## 2. Related Research

Peng et al. [22] put forward the concept of receptive field for the first time through the study of cat visual cortex cells. Based on the concept of the receptive field, Fanthomme and Monasson [23] proposed the neurocognitive machine. The concept of the receptive field also plays an important role in a neural network. Later, the principle of the new cognitive machine is introduced in detail. CNN is built on the basis of a cognitive machine [24]. It is a new model combining Ann and convolution operation. Because of its good versatility, it can be used in the fields of recognition, detection, and tracking. For convolutional neural networks, Lu et al. [25] proposed the lenet-5 model structure based on CNN, which pushed the research of convolutional neural networks to a climax. Falk et al. [26] proposed a new face detection method based on CNN in the literature, which has good robustness to various facial patterns and certain rotation angles and shows that this recognition system does not need expensive preprocessing steps [27]. Subsequently, the convolutional neural network has been developed continuously and has been widely used in various fields and has made a breakthrough. Xiao et al. [28] applied deep learning to the regression task, considered the robustness of the regression algorithm, and analyzed the robustness of the algorithm. According to the relationship between hidden layer and error, a network depth determination method based on reconstruction error was proposed [29].

Principal component analysis (PCA) is a comprehensive statistical method, which transforms complex high-dimensional data into low-dimensional principal components [30]. There is no linear correlation between the extracted principal components, which can reflect most of the information in the original data on the basis of avoiding overlap. With the progress and development of society, the demand and requirement of feature extraction are constantly improved. The method of principal component analysis gradually permeates in all fields of economy and life. In his book, Shang [31] used this method to predict the overall power consumption demand in Australia; Mallord and others [32] used the PCA method to explore the changes of the human force curve and conducted functional principal component analysis on the human force curve to analyze the main factors affecting human fatigue.

## 3. Construction of Model

*3.1. Construction of Recurrent Convolution Neural Network.* The convolutional neural network is a perceptron. The neural network [33–35] uses local connection and weight sharing to reduce the number of parameters and the computational complexity of the network. The simple convolutional neural network is shown in Figure 1; it is a simple convolutional neural network.

Firstly, the input image of the input layer enters the convolution layer, and the convolution operation is carried out through the 3D filter of the convolution layer and the offset vector to generate the feature map; then, a new feature map is obtained by weighted average summation. Then, these

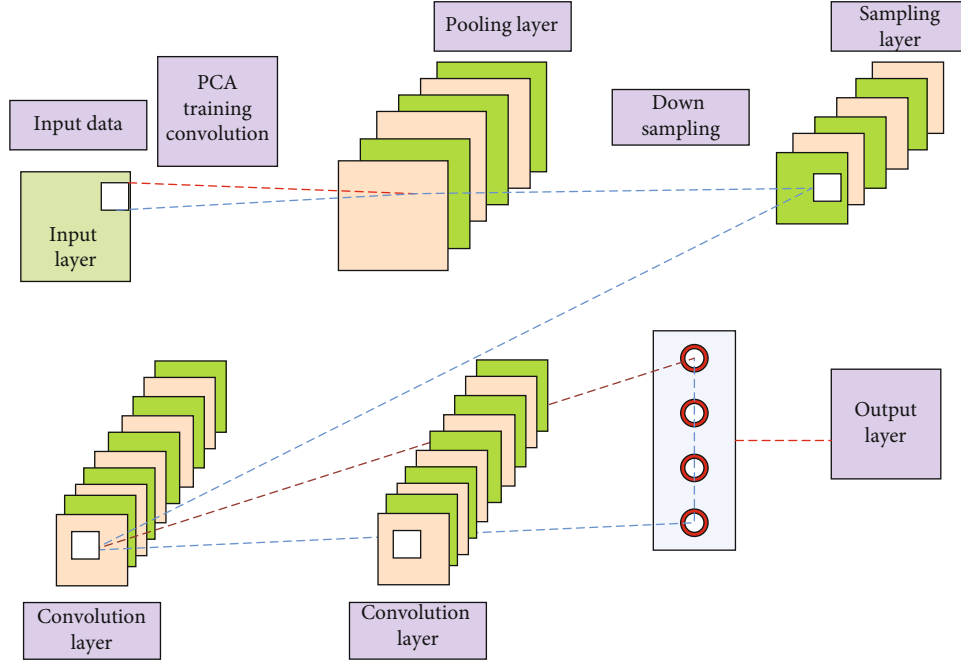


FIGURE 1: Overall structure of convolutional neural network.

feature maps are convoluted by convolution layer filter and output by a pooling layer. Finally, the output layer outputs the classification results through the full connection layer.

The convolution layer mainly implements two key operations: one is local correlation the other is window sliding. Each neuron is regarded as a filter in local correlation, and each filter convolutes the input data. The function of window sliding is to limit each operation in the window when the filter performs convolution operation. When the calculation is completed, the window slides a certain step size for the next convolution operation. The two-dimensional discrete convolution in mathematics satisfies the following formula:

$$X(m, n) = \sum_{u=0}^{v=0} (m \times u) t(u, v) s\left(\frac{m-u}{n-v}\right), \quad (1)$$

where  $x(m, n)$  is the result of convolution,  $s(m, n)$  is the convolution kernel, and  $t(u, v)$  represents the convoluted signal.

Local connection is used to reduce network parameters. For the input two-dimensional image data, each convolution kernel of the convolution layer takes its own template, and after sliding a fixed step distance along the  $x$ -axis and  $y$ -axis of the image, convolution operation is performed to obtain the output response of the corresponding position. When the convolution kernel traverses the current image, the output feature map of the neural element can be obtained. Each convolution layer extracts different features of the input image by setting different convolution kernels, and the parameters of these convolution kernels are updated

by the gradient descent method during network training. The convolution layer is calculated as follows:

$$s_n^p = h\left(\sum_{j=1}^{i=1} s_m^{p-1} + mn\sqrt{f_n^p \times s_j^i}\right), \quad (2)$$

where  $S_n$  is the  $n$ th characteristic graph of convolution layer,  $b_n$  is the set of input characteristic graphs,  $Z$  is the weight matrix of convolution kernel, and  $f$  is the offset.

There are two main purposes in the design of the pooling layer, as shown in Figure 2. When the output size of the build-up layer is  $32 \times 32$ , the size of the pooling layer filter is  $2 \times 2$ . After the pooling layer processing, the output data size is  $16 \times 16$ , and the amount of data is reduced to 1/4 of that before pooling. The reduction of the dimension increases the sparsity of the network, so as to achieve the second purpose, that is, to effectively prevent the network from overfitting. The most important characteristic of the pooling layer is local translation invariance. If the local region of input data is linearly transformed (such as translation and rotation), the output result will not change after downsampling. If we only care about whether a feature appears or not, but not about its position, local translation invariance is very important.

There is a connection layer at the end of the convolutional neural network. Every neuron in the first layer is connected to every neuron in the next layer. The purpose of the full join layer is to map the distribution of the previous layer to the sample label space, then use the function to adjust the relevant process in the learning, and finally give the required prediction. In fact, the fully connected layer is equivalent to the classic multilayer perceptron. Although the full

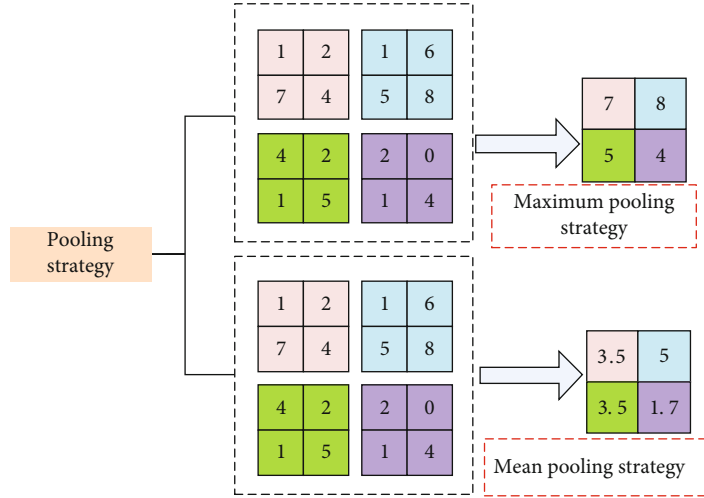


FIGURE 2: Results of different pooling methods.

connection layer is located at the end of the neural network, the parameters of the full connection layer account for most of the parameters of the whole model. Therefore, when designing a convolutional neural network, the number of all connected layers should not be too large.

The recurrent neural network is a kind of network model with forward path and reverse path, which is composed of the input layer, hidden layer, and output layer. The difference between the model and the traditional neural network is that the recurrent neural network can carry out forward and backward propagation. Generally, recurrent neural networks are used to deal with time series problems.

As shown in Figure 3, it is the basic unit of the recurrent neural network. The basic unit of the recurrent neural network is to introduce the characteristic weight matrix  $W$  into the convolutional neural network unit, so as to establish the connection between the last input and the current input.  $x$ ,  $w$ , and  $o$  represent vectors.  $x$  is the value of the input layer.

The basic unit reuse is a recurrent neural network, as shown in Figure 3. The weight matrix  $w$  calculated by each layer of the network will be input into the lower layer network structure.  $t$  represents time,  $x$  represents input,  $o$  represents output, and  $w$  represents the weight matrix of the network, the activation function; the recurrent neural network can be defined as

$$\begin{aligned} w_t &= f \left[ \prod_{t=1} (m \times x) \right] \times w_{t-1}, \\ f_t &= f \left[ \sum_f (h_{t-1}, x_t) \times \sum_t df \right]. \end{aligned} \quad (3)$$

A vector is generated through the activation function to update the data; the output gate processes the output information through the activation function and outputs the final information. As a recurrent neural network, the network is often used for sentence connection and mining the relationship between sample features. This research takes the results

of the students' physical fitness standard test as the research object and makes predictions and analyses.

The network model proposed in this paper consists of Xception, LSTM, and CNN fusion. The innovation lies in the introduction of circular network function and the construction of feature fusion network CNN fusion. Firstly, the sample data are input into the Xception network and LSTM network, respectively, and then, the two weight matrixes obtained by the two calculations are input into the merged layer of CNN fusion at the same time. The features are fused and merged into a weight matrix by this layer, and then, the combined weight matrix is convoluted. Finally, the output layer outputs the category with the highest species probability to complete the classification prediction.

Through the construction of the recurrent convolution network model, the student physical health standard test is realized, and the flow is shown in Figure 4. Before the initialization of the network model, the original data is denoised, enhanced, rotated, and clipped for the first time. Then, the correlation between image features is extracted by Xception and LSTM, respectively, and the weight matrix of the two networks is output to CNN fusion, and the shape tensor of the weight matrix is multiplied by each element for fusion learning.

**3.2. Model Based on Recurrent Convolution Neural Network and Principal Component Analysis.** In the training process of traditional CNN, the initialization of weight is randomly generated. Although a better convergence effect can be obtained through the gradient descent method, the convergence direction of the mean square error cannot be found quickly when the iteration starts or the number of iterations is limited, which affects the convergence speed and ultimately the recognition rate. Based on the CNN model and PCA feature extraction method, this paper realizes the CNN data analysis method of multilayer convolution kernel parameter initialization. Through principal component analysis of different layers in the network, the algorithm obtains the eigenvector that can represent the corresponding layer and takes the

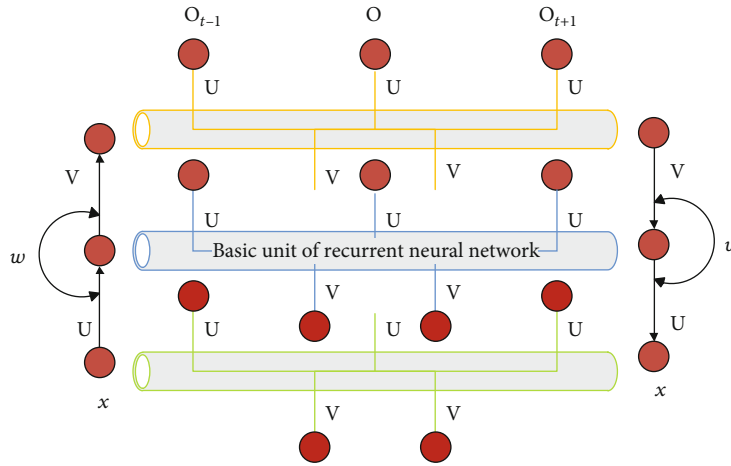


FIGURE 3: Node structure of recurrent neural network.

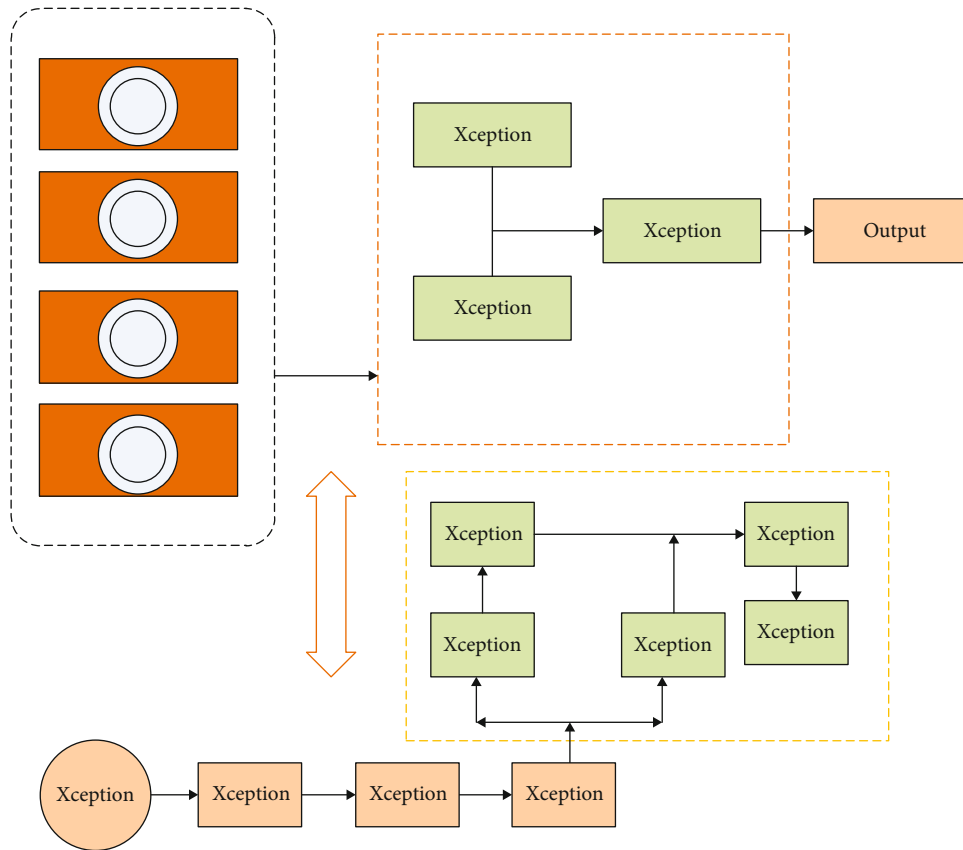


FIGURE 4: Construction of recurrent convolution network.

eigenvector as the initialization value of the convolution kernel of the corresponding convolution layer, so as to avoid the defects of many iterations and low recognition rate in the random initialization of the convolution kernel. The flow chart is shown in Figure 5. Firstly, the training data and test data are preprocessed, the symbols in the network connection data are converted into numbers, and each data is marked in advance. After PCA dimensionality reduction, it is trained in the designed CNN network. After many times of training, according to the

comparison of test set results, the analysis model with the best performance is obtained.

#### 4. Principal Component Analysis of Health Standard Test Results

4.1. General Process of Principal Component Analysis. Principal component analysis, like traditional factor analysis, is a data processing method to reduce dimension and seek commonness among samples. In the processing of image, voice,

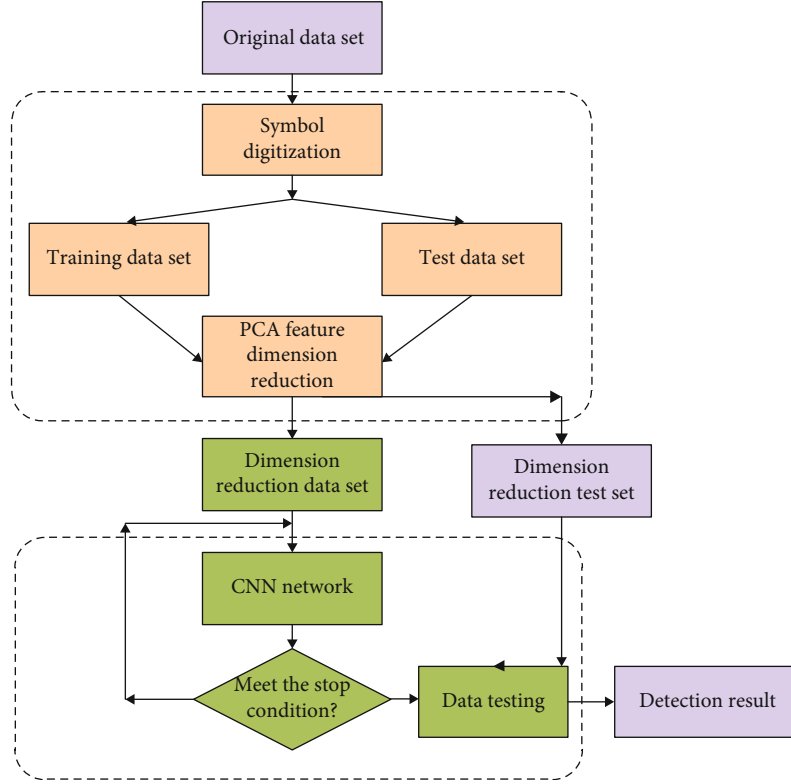


FIGURE 5: Detection flow chart based on PCA-CNN.

and other data problems, we usually use this method when we need to reduce the dimension and extract the corresponding feature commonness or when the attribute dimension is higher, and we need to use another lower dimension expression. Dimension reduction is very necessary in data processing. Firstly, there is a correlation between variables, that is, multicollinearity, which will lead to the instability of solution space, and the result may be incoherent. Secondly, high-dimensional space is sparse. Thirdly, too many variables hinder the establishment of rules. In data processing such as machine learning, too much data may cause overfitting or underfitting, which can better predict and classify problems. Finally, there are many characteristics of a thing or problem, but the variables representing the characteristics may have internal information overlap, so it is difficult to consider the potential relationship between variables in the analysis of variables. Based on the above reasons, we need to reduce the dimension of data for three purposes: First, the most intuitive is to reduce the number of attributes. Secondly, the dimensionality reduction can ensure that the variables are independent of each other. Finally, it is convenient to explain the meaning of the calculated components. The purpose of PCA is to change the properties of data feature variables, so all operations are based on feature dimension.

For the original data set,  $x$  is the number of samples and  $y$  is the dimension of samples. Through

$$c_{ij} = d_{ij} \times c - \frac{\sum_{k=1}^x d_{kj}}{m}. \quad (4)$$

Centralize the data, get the corresponding matrix, use the matrix to calculate the covariance matrix  $C$ , and finally calculate the eigenvalue and eigenvector of  $C$ . As for the last step of eigenvalue selection rules, according to their different needs, the selection method is different. In data visualization, the first two largest eigenvalues and corresponding eigenvectors are usually selected. We can get two-dimensional data by mapping, which is easy to display by graphics. Another way is to score: assume that there are eigenvalues, according to

$$u = \frac{\sum_{i=0}^s \beta_s \times \beta_j}{\sum_{j=0}^t \beta_j}. \quad (5)$$

Among them, the denominator is the sum of all eigenvalues, and the numerator is the sum of the top  $s$  largest eigenvalues, which is the score rate. When we set a threshold, as long as the score rate exceeds this value, then these eigenvectors are the eigenvectors we need, namely, pattern vectors.

*4.2. Application of Principal Component Analysis in Comprehensive Evaluation of College Students' Physical Health.* According to the requirements of standard documents, students' physical fitness test items mainly include three aspects: physical fitness, body shape, and body function. In terms of body shape, it mainly tests body mass index; that is, it evaluates the body's symmetry, reflecting the body's growth and development level and nutritional

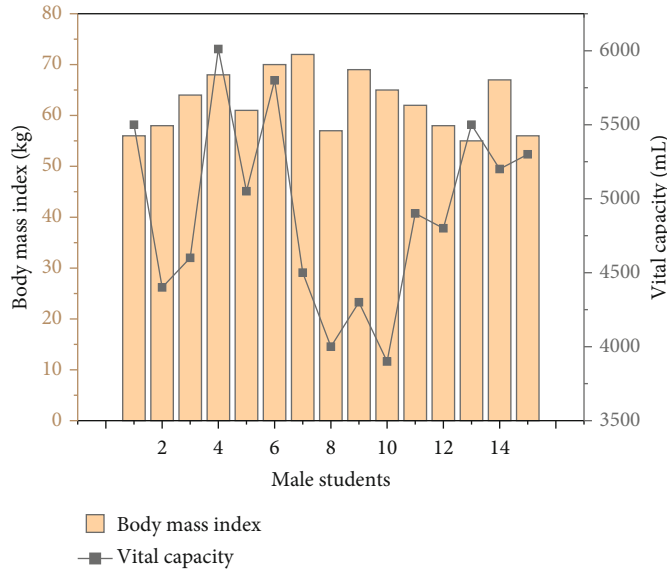


FIGURE 6: Body and shape data of male students.

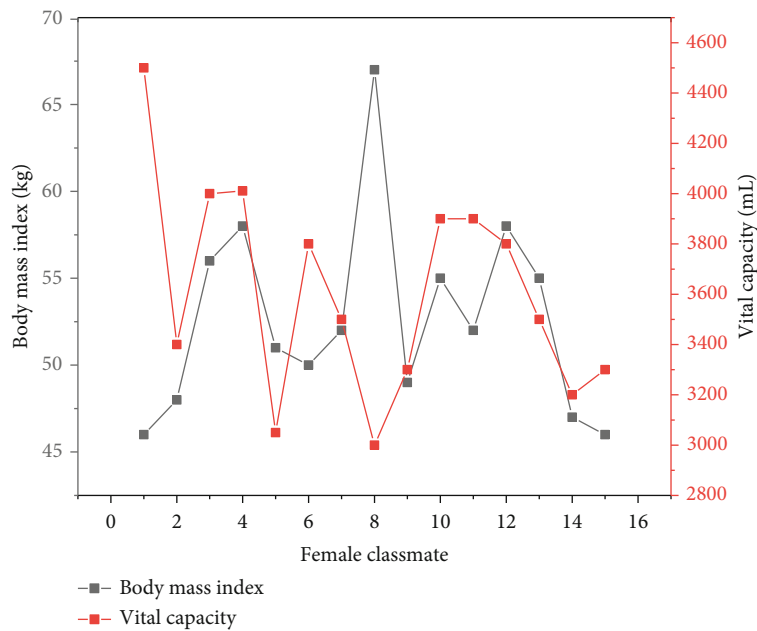


FIGURE 7: Body and shape data of female students.

status. In terms of physical function, it mainly tests vital capacity; that is, it evaluates the maximum ventilation capacity of human breathing, reflecting the volume and expansion capacity of the lung. In terms of physical fitness, it mainly tests the standing long jump, sitting forward bending, 50 m running, pull-up, and 1000 m long run. Standing long jump evaluates the long jump ability of the human body, reflecting the explosive ability of human lower limbs and the coordination ability of the human body. The forward flexion of the sitting body evaluates the range of motion that can be achieved by the trunk, waist, hip, and other joints of the human body in a static state and reflects the development

level of physical flexibility. The speed quality of the human body is evaluated in 50 m race, which reflects the explosive power, sensitivity, reaction, and flexibility of the human body. The upper limb overhanging strength, shoulder strap strength, and grip strength were measured by the pull-up test to reflect the upper limb strength endurance level. The aerobic and anaerobic endurance of the human body is evaluated by 1000 m long-distance running, which reflects the cardiopulmonary function and endurance level of the human body.

On the basis of theoretical research, combined with the actual data of students' physical fitness test, the body shape



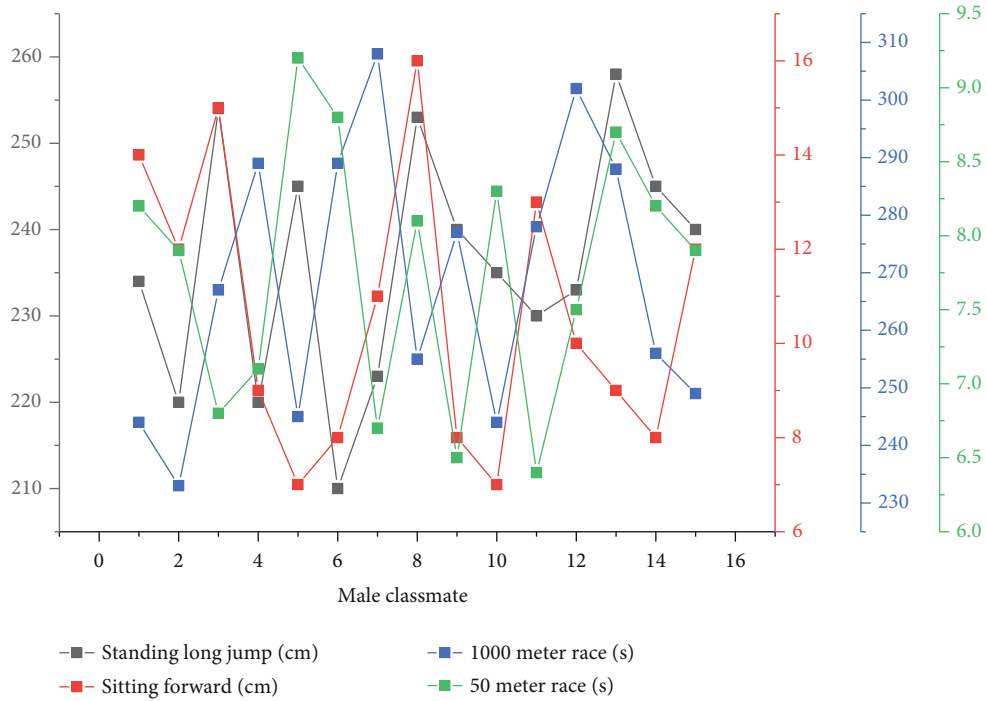


FIGURE 8: Data chart of male students' comprehensive physical fitness.

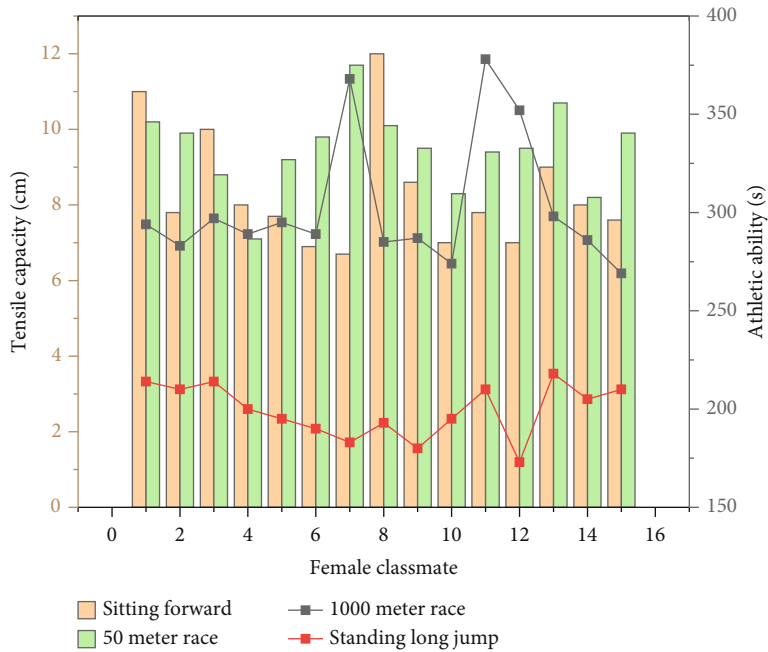


FIGURE 9: Data chart of female students' comprehensive physical fitness.

(body mass index), physical function (vital capacity), and physical fitness (standing long jump, sitting forward, 50 m running, pull-up, and 1000 m running) are taken as the comprehensive evaluation indexes of students' physical health level. Taking 200 boys' and girls' physical health test data as an example, this paper explains the specific application of principal component analysis in the comprehensive eval-

uation of physical health. The physique test data of 15 boys and girls were randomly selected, as shown in Figures 6–9.

As shown in Figure 6, the body mass index of 25 students is basically about 50-60 kg, but their vital capacity is quite different. Through qualitative investigation of the six evaluation indexes reflecting the physical health of college students, it is obvious that there is a relatively strong

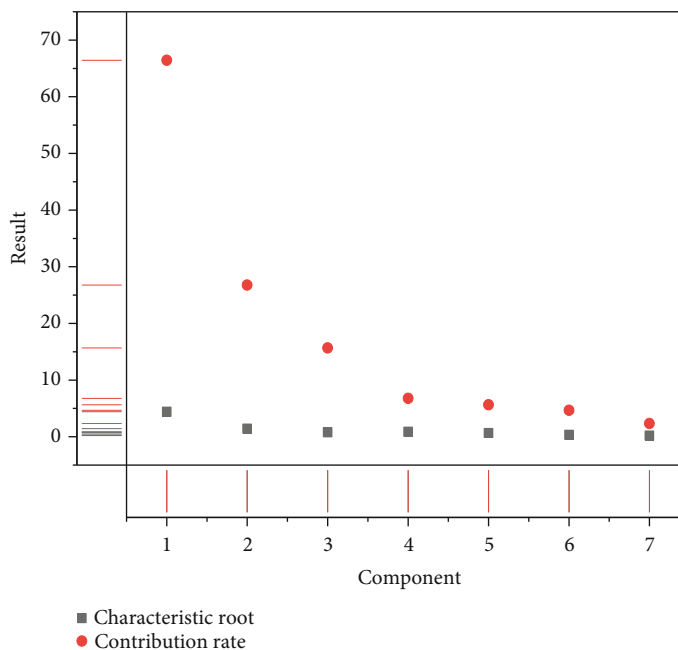


FIGURE 10: Principal component analysis.

correlation between some indexes; for example, there may be a strong correlation between standing long jump and sitting forward bending. There may be a strong correlation among vital capacity, 50 m running, and 1000 m running. In order to test these possible situations, SPSS software is used to analyze the correlation of these seven indicators, and it is found that there is a strong correlation between some indicators. In order to avoid information overlapping and affect the objectivity of the final comprehensive evaluation results, it is necessary to use principal component analysis to transform these seven indicators into several comprehensive indicators with low correlation. SPSS software is used to analyze the seven indicators, and the eigenvalues and contribution rate of the corresponding correlation coefficient matrix are obtained (Figure 10).

It can be seen from Figure 10 that the eigenvalues of the first three components are significantly higher than the other components, and the eigenvalues of the other components are relatively small. The first three principal components are selected as principal components for analysis, and the cumulative contribution rate of the eigenvalues of the first three principal components reaches 90%, indicating that the principal component analysis method is very effective in this problem. From the corresponding coefficients of each principal component, it can be seen that the first principal component contains information about standing long jump, sitting forward, pull-up, 50 m run, and 1000 m run; the second principal component contains vital capacity information; the third principal component contains weight index information. Substituting the standardized data of each student's original indicator into the expression of the three principal components, the values of the three principal components corresponding to each student are obtained. Then, the contribution rates corresponding to the three principal

components are used as their respective weights, and a comprehensive evaluation model of college students' physical fitness based on principal component analysis is obtained.

## 5. Conclusion

The principal component analysis and prediction of the test results of students' physical fitness standards based on the convolution regression neural network are studied. By constructing a recurrent neural network and a convolutional neural network, the principal component analysis method is used to analyze the students' body mass index, vital capacity, standing long jump, sitting and forward bending, 50 m running, pull-ups, and 1000 m running, and three principal components are found more reasonable; the first principal component has the highest contribution rate, which mainly reflects the information of the five indicators of standing long jump, sitting forward bend, pull-up, 50 m running, and 1000 m running; that is, physical fitness indicators have a greater impact on the health of college students. This is in line with the current status of the physical health of college students, indicating that the results of the principal component analysis are scientific and reasonable.

## Data Availability

The data used to support the findings of this study are included within the article.

## Conflicts of Interest

The author does not have any possible conflicts of interest.

## Acknowledgments

This study was supported by the University-Industry Collaborative Education Program (201802088040).

## References

- [1] Y. M. Zhang, T. S. Lee, M. Li, F. Liu, and S. Tang, "Convolutional neural network models of V1 responses to complex patterns," *Journal of Computational Neuroscience*, vol. 46, no. 1, pp. 33–54, 2019.
- [2] S. Qi, X. Ning, G. Yang et al., "Review of multi-view 3D object recognition methods based on deep learning," *Displays*, vol. 69, p. 102053, 2021.
- [3] S. A. Rahman and D. A. Adjeroh, "Deep learning using convolutional LSTM estimates biological age from physical activity," *Scientific Reports*, vol. 9, no. 1, pp. 1–15, 2019.
- [4] G. Manogaran, P. M. Shakeel, H. Fouad et al., "Wearable IoT smart-log patch: an edge computing-based Bayesian deep learning network system for multi access physical monitoring system," *Sensors*, vol. 19, no. 13, p. 3030, 2019.
- [5] Y. Jiang, X. Gu, D. Wu, W. Hang, J. Xue, and S. Qiu, "A novel negative-transfer-resistant fuzzy clustering model with a shared cross-domain transfer latent space and its application to brain CT image segmentation," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 18, no. 1, pp. 40–52, 2020.
- [6] J. Zhang, Y. Liu, H. Liu, and J. Wang, "Learning local-global multiple correlation filters for robust visual tracking with Kalman filter redetection," *Sensors*, vol. 21, no. 4, p. 1129, 2021.
- [7] L. Zhang, X. Wang, X. Dong, L. Sun, W. Cai, and X. Ning, "Finger vein image enhancement based on guided tri-Gaussian filters," *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 1, pp. 17–23, 2021.
- [8] C. Xie and A. Kumar, "Finger vein identification using convolutional neural network and supervised discrete hashing," *Pattern Recognition Letters*, vol. 119, pp. 148–156, 2019.
- [9] C. Ying, Y. Shuyu, L. Jing, D. Lin, and Q. Qi, "Errors of machine translation of terminology in the patent text from English into Chinese," *ASP Transactions on Computers*, vol. 1, no. 1, pp. 12–17, 2021.
- [10] W. J. Lee, L. N. Peng, S. T. Chiou, and L. K. Chen, "Physical health indicators improve prediction of cardiovascular and all-cause mortality among middle-aged and older people: a national population-based study," *Scientific Reports*, vol. 7, no. 1, pp. 1–8, 2017.
- [11] C. Miles, A. Bohrdt, R. Wu et al., "Correlator convolutional neural networks as an interpretable architecture for image-like quantum matter data," *Nature Communications*, vol. 12, no. 1, p. 3905, 2021.
- [12] H. Raat, G. J. Bonsel, M. L. Essink-Bot, J. M. Landgraf, and R. J. Gemke, "Reliability and validity of comprehensive health status measures in children: the Child Health Questionnaire in relation to the Health Utilities Index," *Journal of Clinical Epidemiology*, vol. 55, no. 1, pp. 67–76, 2002.
- [13] C. M. Chiang and C. M. Lai, "A study on the comprehensive indicator of indoor environment assessment for occupants' health in Taiwan," *Building and Environment*, vol. 37, no. 4, pp. 387–392, 2002.
- [14] C. W. Coley, W. Jin, L. Rogers et al., "A graph-convolutional neural network model for the prediction of chemical reactivity," *Chemical Science*, vol. 10, no. 2, pp. 370–377, 2019.
- [15] H. Wang, S. Ding, D. Wu, Y. Zhang, and S. Yang, "Smart connected electronic gastroscope system for gastric cancer screening using multi-column convolutional neural networks," *International Journal of Production Research*, vol. 57, no. 21, pp. 6795–6806, 2019.
- [16] A. N. Gorban, E. M. Mirkes, and I. Y. Tyukin, "How deep should be the depth of convolutional neural networks: a backyard dog case study," *Cognitive Computation*, vol. 12, no. 2, pp. 388–397, 2020.
- [17] K. Gong, J. Guan, K. Kim et al., "Iterative PET image reconstruction using convolutional neural network representation," *IEEE Transactions on Medical Imaging*, vol. 38, no. 3, pp. 675–685, 2019.
- [18] B. Hrnjica and O. Bonacci, "Lake level prediction using feed forward and recurrent neural networks," *Water Resources Management*, vol. 33, no. 7, pp. 2471–2484, 2019.
- [19] I. Obeso, D. Ledbetter, M. Aczon, E. Laksana, M. Wintner, and R. Wetzel, "694: continuous risk of infection prediction using recurrent neural networks in a pediatric ICU," *Critical Care Medicine*, vol. 49, no. 1, pp. 342–342, 2021.
- [20] R. Laatar, C. Aloulou, and L. H. Belguith, "Disambiguating Arabic words according to their historical appearance in the document based on recurrent neural networks," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 19, no. 6, pp. 1–16, 2020.
- [21] J. Villegas, K. Markov, J. Perkins, and S. J. Lee, "Prediction of creaky speech by recurrent neural networks using psycho-acoustic roughness," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 2, pp. 355–366, 2020.
- [22] P. A. Peng, W. Zhang, Y. Zhang, Y. Xu, H. Wang, and H. Zhang, "Cost sensitive active learning using bidirectional gated recurrent neural networks for imbalanced fault diagnosis," *Neurocomputing*, vol. 407, pp. 232–245, 2020.
- [23] A. Fanthomme and R. Monasson, "Low-dimensional manifolds support multiplexed integrations in recurrent neural networks," *Neural Computation*, vol. 33, no. 4, pp. 1063–1112, 2021.
- [24] A. Ali, B. M. Margetts, and A. A. Zainuddin, "Exploration of the principal component analysis (PCA) approach in synthesizing the diet quality of the Malaysian population," *Nutrients*, vol. 13, no. 1, p. 70, 2021.
- [25] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan, "Tensor robust principal component analysis with a new tensor nuclear norm," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 4, pp. 925–938, 2020.
- [26] A. R. Falk, J. C. Lamsdell, and E. Gong, "Principal component analysis of avian hind limb and foot morphometrics and the relationship between ecology and phylogeny," *Paleobiology*, vol. 47, no. 2, pp. 314–336, 2021.
- [27] Y. Yu, "Functional principal component analysis: a robust method for time-series phenotypic data," *Plant Physiology*, vol. 183, no. 4, pp. 1422–1423, 2020.
- [28] W. Xiao, X. Huang, F. He, J. Silva, S. Emrani, and A. Chaudhuri, "Online robust principal component analysis with change point detection," *IEEE Transactions on Multimedia*, vol. 22, no. 1, pp. 59–68, 2020.
- [29] A. L. Machidon, F. del Frate, M. Picchiani, O. M. Machidon, and P. L. Ogrutan, "Geometrical approximated principal component analysis for hyperspectral image analysis," *Remote Sensing*, vol. 12, no. 11, p. 1698, 2020.
- [30] A. Djerida, Z. Zhao, and J. Zhao, "Background subtraction in dynamic scenes using the dynamic principal component analysis," *IET Image Processing*, vol. 14, no. 2, pp. 245–255, 2020.

- [31] A. T. de Souza, L. A. T. X. Carneiro, O. P. da Silva Junior, S. L. de Carvalho, and J. H. P. Américo-Pinheiro, "Assessment of water quality using principal component analysis: a case study of the Marrecas stream basin in Brazil," *Environmental Technology*, vol. 42, no. 14, pp. 1–21, 2020.
- [32] X. Yue and H. Zhang, "Grasshopper optimization algorithm with principal component analysis for global optimization," *The Journal of Supercomputing*, vol. 76, no. 3, pp. 1–27, 2020.
- [33] Y. Tong, L. Yu, S. Li, J. Liu, H. Qin, and W. Li, "Polynomial fitting algorithm based on neural network," *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 1, pp. 32–39, 2021.
- [34] P. Chen, W. Li, L. Sun, X. Ning, L. Yu, and L. Zhang, "LGCN: learnable Gabor convolution network for human gender recognition in the wild," *IEICE Transactions on Information and Systems*, vol. 102, no. 10, pp. 2067–2071, 2019.
- [35] Z. Chu, M. Hu, and X. Chen, "Robotic grasp detection using a novel two-stage approach," *ASP Transactions on Internet of Things*, vol. 1, no. 1, pp. 19–29, 2021.

## Research Article

# Influencing Factors of Athletes' Injury Rehabilitation from the Perspective of Internal Environment

Xiang Huang<sup>1</sup> and Xiaoping Wang<sup>2</sup> 

<sup>1</sup>School of Physical Education and Health, Yulin Normal University, Yulin 537000, China

<sup>2</sup>School of Biology and Pharmacy, Yulin Normal University, Yulin 537000, China

Correspondence should be addressed to Xiaoping Wang; 201410730212@stu.shmtu.edu.cn

Received 23 July 2021; Revised 8 August 2021; Accepted 9 August 2021; Published 2 September 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Xiang Huang and Xiaoping Wang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Athlete's injury recovery is related to the athlete's personal value. A scientific and effective rehabilitation program will help athletes overcome their illnesses and return to the game as soon as possible. Based on the literature review and the internal environment perspective, this paper constructs a model of factors affecting athletes' injury rehabilitation. Through the empirical analysis of 129 questionnaires, we have verified the research hypothesis of each factor. The research results show that psychological adjustment, rehabilitation learning, and video reflection have a significant positive impact on athletes' injury rehabilitation, while imagery has no significant impact. This research provides a reference plan for athletes to adopt effective injury rehabilitation training methods. At the same time, we have also enriched the research literature on athletes' injury rehabilitation solutions.

## 1. Introduction

A healthy body is a key element for athletes to maintain their competitive advantage [1]. However, once an athlete is injured and fails to recover smoothly, it may affect the athlete's entire career [2]. Basketball player McGrady and hurdler Liu Xiang are typical examples. These athletes were plagued by injuries, and due to the poor recovery of the injury in the later period, the sports career came to an abrupt end. In addition, short-term injuries and poor recovery may also reduce the commercial value of athletes. For example, due to major athletes' injuries, NBA players may not be able to obtain a "big contract" during the health period when trading.

Sports injury rehabilitation [3] refers to comprehensive measures taken against dysfunction caused by athletes' injuries (e.g., bones, joints, muscles, and ligaments) [4]. The purpose is to improve and enhance the athletic function of athletes, so that patients can return to society [4]. At present, the main rehabilitation measures for athletes' injuries include physical therapy (e.g., exercise therapy and physical

therapy), occupational therapy, and the application of orthotics and assistive devices [5]. Due to the huge consequences of athletes' injuries, various countries and teams are paying more and more attention to the rehabilitation and prevention of injuries.

Many scholars have carried out relevant research on specific measures for athletes' injury rehabilitation. For example, Hare et al. [1] and Sordoni et al. [4] studied the influence of intention on athletes' injury rehabilitation. Forsdyke et al. [2] analyzed the influence of psychological factors on athletes' injury rehabilitation through literature review; King et al. [6] put forward that authorization, participation, providing feedback, and transparency are four important habits to improve the effectiveness of athletes' injury rehabilitation. It can be seen that there are various measures for athletes' injury rehabilitation, and this is a process that is affected by both internal and external environments. In order to enhance the accuracy of the research, we only discuss the question of athletes' injury rehabilitation from an internal perspective in this paper. This paper have answered the question of what factors affect

the rehabilitation of athletes' injury from the internal perspective. The research data of this paper was collected from the questionnaire and is analyzed by the SmartPLS software. Our study has enriched the related literature on athletes' injury rehabilitation; in particular, based on the internal perspective, we examined the effects of psychological adjustment, imagery, rehabilitation learning, and learning from the video on athletes' injury rehabilitation. In addition, this paper provided a feasible reference plan for athletes' injury rehabilitation.

The other parts of this research are arranged as follows: The second part is a literature review; the third part designs the theoretical framework of the research and puts forward the research hypothesis; the fourth part is the research methods and empirical results; the fifth part is the conclusion.

## 2. Literature Review

*2.1. Athletic Injuries.* Athletic injuries, as a common exercise disease, refer to various injuries that occur during exercise [7]. Generally speaking, the location of sports injury is related to sports events and specific technical characteristics [8]. For example, the injured parts of basketball players are mostly feet, wrists, shoulders, and waist. This is because there are more technical movements such as jumping, running, reclining, shooting, and accelerating in basketball [9].

There are many reasons for athletes' injuries, but they can be roughly divided into five categories [10, 11]: (1) insufficient training level, (2) irregular exercises, (3) failure to do preexercise or insufficient preexercise, (4) unintentional injury of other athletes, and (5) improper organization of competition work. Typical symptoms of athletic injuries include pain, swelling, and limited movement [10]. Athletic injuries can be identified by X-ray, magnetic resonance imaging, or CT [8, 11].

Once a sports injury occurs, it needs to be checked and treated in time [12]. Because athletes' injuries are more acute than chronic, improper or untimely treatment of acute injuries may miss the best opportunity for injury treatment on the one hand, and on the other hand, it may also lead to the conversion of acute injuries to chronic injuries [13]. Therefore, athletes' injuries need to be taken seriously by athletes, team staff, and sports delegations from all over the world.

*2.2. Measures for Injury Rehabilitation.* Sports injury rehabilitation refers to the treatment measures taken after injury [14], which is fundamentally different from injury prevention. Due to the different parts of the body injured by athletes, there are also differences in specific rehabilitation measures [15]. Due to research needs, this article mainly discusses universal rehabilitation measures and their influencing factors based on the internal environment perspective.

The internal environment perspective [16], in which specialized rehabilitation strategies focused on the athletes themselves are discussed. For example, Gledhill and Ivarsson [17] proposed that athletes' injuries can cause athletes to experience psychosocial stress, anxiety, and emotional dis-

stress. Furthermore, their ability to make rational decisions about rehabilitation activities may also be affected. Therefore, psychosocial factors have an important influence on the rehabilitation of athletes [18]. Forsdyke [19] proposed that imagery has a positive effect on sports injury rehabilitation. Specifically, imagination can help improve the patient's injury experience during the rehabilitation process. In addition, deep learning [20–22] and vision-based technologies [23–25] are also widely used in the field of sports injury rehabilitation research.

The above rehabilitation measures for athletes' injuries are all based on the athlete's internal perspective. It can be seen that based on the perspective of the internal environment, scholars have provided a wealth of literature on sports injury rehabilitation. However, relatively speaking, what is the effect of these internal measures on the rehabilitation of athletes, and what is the relationship between them? These issues need to be discussed in depth.

## 3. Theoretical Framework and Research Hypothesis

*3.1. Theoretical Framework.* Based on the above literature analysis, this paper constructs a model of factors affecting athletes' injury rehabilitation based on the internal environment perspective. See Figure 1 for details. The theoretical model mainly includes 4 internal factors, namely, psychological adjustment, imagery, rehabilitation learning, and video reflection. The two factors of psychological adjustment and imagination mainly come from previous studies. Rehabilitation learning and learning from the video are proposed in this article. Regarding the specific role of the latter two factors, we will elaborate in the research hypothesis section.

### 3.2. Research Hypothesis

*3.2.1. Psychological Adjustment.* Psychosocial factors are increasingly recognized as important factors affecting the rehabilitation and treatment of athletes [26] because, in sports, psychological factors are as important as the physical ability and talent of athletes. Especially after athletes' injuries, athletes need to release the pressure through psychological adjustment, so as to maintain an optimistic and positive attitude [27]. The study of Forsdyke et al. [2] found that athletes' emotions have a clear relationship with rehabilitation effects, and injuries, athletes' fear, anxiety, and self-confidence are related to rehabilitation results. The research of Gledhill and Ivarsson [17] also shows that athletes' injuries often cause athletes' psychological stress, anxiety, and emotional distress. If these psychological problems are not handled well, they may affect their recovery process. It can be seen that psychological adjustment has a greater impact on athletes' injury rehabilitation. Ahern and Lohr [26] believe that athletes have good psychological adjustment capabilities, which will help their recovery from later injuries. Based on the above analysis, we put forward the following hypotheses:

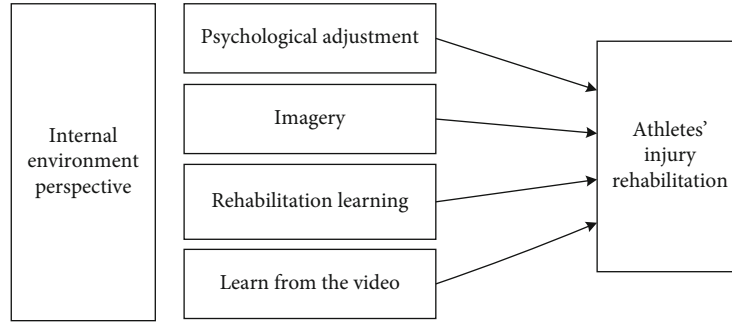


FIGURE 1: Research framework.

TABLE 1: Scale.

Latent variables	Observed variables	References
Psychological adjustment (PA)	1a: I think psychological adjustment is very important for injury recovery. 1b: I think good psychological adjustment can help me overcome pain. 1c: I think that recovery from injury cannot be separated from self-adjustment.	Gledhill and Ivarsson [17]; Ahern and Lohr [26]
Imagery (I)	2a: I often use imagery in the course of injury rehabilitation. 2b: I think the use of image therapy to treat pain can improve our satisfaction with recovery. 2c: I think image therapy can be used for pain management.	Hare et al. [1]; Law et al. [29]
Rehabilitation learning (RL)	3a: I often learn rehabilitation-related skills during the injury rehabilitation period. 3b: I often learn about injury prevention during the injury recovery period. 3c: I often learn theories related to improving the professional level of sports during the injury recovery period.	Jespersen [30]; Eastman and Chang [31]
Learn from the video (LftV)	4a: I often watch my own sports videos during the injury recovery period to reflect and learn. 4b: During my injury rehabilitation period, I often watch sports videos of outstanding athletes in the same field to reflect and learn. 4c: I think learning from the video will help me through the difficult rehabilitation period.	Wang and Parameswaran [32]
Athletes' injury rehabilitation (AIR)	a: I think the process of injury recovery is affected by many factors. b: I think the key to good injury recovery lies in having a scientific coping style. c: I think the injury recovery period is not only a recovery process but also a good learning process.	Dhillon et al. [5]

H1: The stronger the athlete's psychological adjustment ability, the more helpful it is to recover from athletes' injuries.

**3.2.2. Imagery.** Imagery is widely used in practical treatment as a scientific rehabilitation method for athletes' injuries [4]. Green [28] studied the influence of imagery on sports injury rehabilitation earlier, and he demonstrated the application of imagery techniques in the chronological sequence of athletes' injuries. His research found that describing standards related to performance, athletes can use it to compare their progress in the recovery process [28]. The study of Hare et al. [1] shows that imagery ability has an impact on athletes' perception of injury response and imagination use function. In addition, a study by Law et al. [29] found that athletes use imagery for pain management, outdoor training, and competition environments. Athletes' use of image

therapy to treat pain can increase their satisfaction with recovery. Accordingly, we propose the following research hypotheses:

H2: Athletes use image therapy to help their recovery from athletes' injuries.

**3.2.3. Rehabilitation Learning.** Rehabilitation learning refers to learning skills related to rehabilitation, knowledge related to injury prevention, and theories related to improving the professional level of sports after an athlete's injury [30]. During the rehabilitation process, it is very important for athletes to maintain continuous learning. Athletes should shift from passive rehabilitation to active rehabilitation from injury. In this way, athletes are required to carry out necessary rehabilitation studies [31]. Learning skills related to rehabilitation can help athletes quickly adapt to their rehabilitation life; learning knowledge related to injury prevention

TABLE 2: Descriptive statistical results of the survey.

Basic information	Items	Rate
Age	18-25	68%
	26-30	29%
	>30	3%
Gender	Female	41%
	Male	59%
Whether they are recovering from injury	Yes	100%
	No	0%
Sports field	Track and field	17%
	Ball	35%
	Swimming	8%
	Other	40%
Education level	Junior college or below	6%
	Bachelor's degree	73%
	Master degree or above	21%

can help athletes avoid reinjury in the later period and reduce the probability of later injury of athletes; it assists players in entering the training environment in advance and preparing for their comeback after recovery by learning theories connected to enhancing the professional level of sports. Based on the above analysis, we propose the following hypotheses:

H3: Athletes use rehabilitation learning therapy to help their recovery from athletes' injuries.

**3.2.4. Video Reflection.** Video reflection refers to athletes who reflect and learn by watching their own sports videos and sports videos of outstanding athletes in the same field during the recovery from injury. The methods of analyzing and learning videos can be seen in many sports events [32]. For example, before a basketball game, many NBA teams will broadcast the opponent's game video to analyze the characteristics of the opponent's players and team tactics. Athletes watch their own sports videos during their rehabilitation. On the one hand, they can find improper actions that can easily lead to injuries from the videos. On the other hand, they can help maintain a good competitive state. And learn the sports videos of outstanding athletes in the same field, you can compare and analyze your own technical actions and then summarize the subsequent technical actions that can be improved. Accordingly, we propose the following research hypotheses:

H4: Athletes use video-retrospective therapy to help their recovery from athletes' injuries.

## 4. Research Methods and Results

**4.1. Survey Design.** In order to identify the factors affecting athletes' injury rehabilitation based on the internal environment perspective, we conducted a questionnaire design based on research hypotheses. The questionnaire mainly includes two parts. The first part is the basic information of the surveyed person (injured athletes who are in the

TABLE 3: Results of reliability and validity measurement.

Construct	CA	CR	AVE
PA	0.758	0.823	0.642
I	0.713	0.785	0.733
RL	0.765	0.812	0.657
LfV	0.834	0.877	0.769
AIR	0.791	0.836	0.699

recovery stage), including age, gender, education level, sports engaged in, and whether they are in the injury recovery stage. Meanwhile, the basic information in the first part is the control variables of this study. The second part is the four research variables (mental adjustment, imagery, rehabilitation learning, and video reflection) and one target variable (the athlete's injury rehabilitation) proposed in the previous article. In the specific operation, we use the 5-point Likert scale to quantify the observed variables of each latent variable [33, 34]. On the basis of referring to previous studies, we developed the scale shown in Table 1.

**4.2. Data Collection.** In order to ensure the validity of the questionnaire, we invited 4 athletes in the rehabilitation period to conduct a pretest of the questionnaire before the formal survey. At the same time, we have corrected some ambiguous items based on their feedback. Our questionnaire is mainly conducted online. The investigation will start in May 2021 and will be completed in July 2021. A total of 165 questionnaires were returned in this survey, of which 129 were valid questionnaires. The effective rate of questionnaires returned was 78.2%. The survey was conducted anonymously, and all the information of the respondents was only used for scientific research. The descriptive statistical results of this survey are shown in Table 2.

**4.3. Data Analysis.** This paper uses the SmartPLS software and PLS-SEM data modeling [35] to conduct empirical



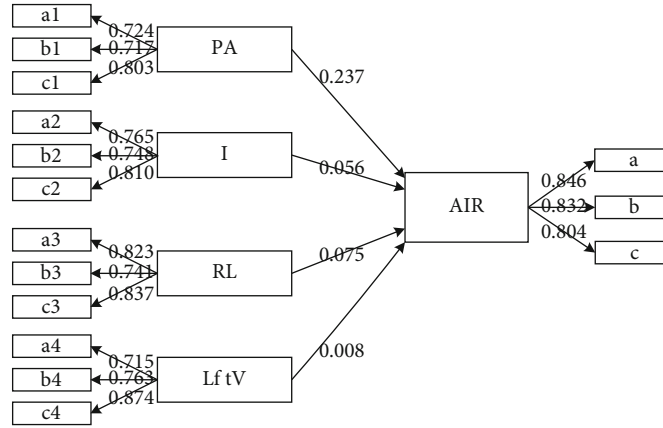


FIGURE 2: Model results.

research on the factors affecting athletes' injury rehabilitation. This research method has a very wide range of applications in the analysis of influencing factors. For example, Thaker et al. [36] used the SmartPLS software to study the factors that affect social media advertising. Silaparasetti et al. [37] used the SmartPLS software to analyze the impact of occupational health and safety on workers' behavior. It can be seen that this method is suitable for this study.

**4.3.1. Reliability and Validity of Measurement.** This paper analyzes the reliability and validity of the research scale and model. The results (see Table 3) show that the relevant indicators (Cronbach's alpha (CA), combined reliability (CR), and average extraction variance (AVE)) have reached the corresponding threshold [38]. In addition, the external load value of each latent variable is greater than 0.7 (see Figure 2 for details). It can be determined that the scale and model of this study have good reliability and validity [38]. At the same time, it can be seen from the data analysis results in Table 4 that there is no multicollinearity among the latent variables [39]. The value of  $R^2$  in the model is greater than 0.5, indicating that the model constructed in this study has good explanatory power [39]. In addition, the value of the model fit in this paper is 0.035, less than 0.06, indicating that the model constructed has good adaptability [38].

**4.3.2. Hypothesis Testing.** This paper uses the SmartPLS software to verify the research hypothesis. The results are shown in Table 5. The research results show that hypotheses H1, H3, and H4 passed the hypothesis verification ( $P$  value is significant), while H2 did not pass the hypothesis verification ( $P$  value is not significant) [39]. From this, we can see that psychological adjustment, rehabilitation learning, and learn from the video have a significant positive impact on athletes' injury rehabilitation, while imagery has no significant impact. By comparison, we found that psychological adjustment, rehabilitation learning, and learn from the video support the majority of previous conclusions. However, our results suggest that imagery has no significant effect on athletes' injuries rehabilitation, which is contrary to the results of Hare et al. [1], Green [28], and Law et al. [29].

TABLE 4: Average Variance Extracted (AVE) square root and factor correlation coefficient.

	H	H1	H2	H3	H4
AIR	<b>0.947</b>				
PA	0.601	<b>0.866</b>			
I	0.426	0.576	<b>0.787</b>		
RL	0.385	0.486	0.614	<b>0.704</b>	
LftV	0.233	0.358	0.332	0.431	<b>0.692</b>

Diagonal elements are the square root of Average Variance Extracted (AVE).

TABLE 5: Results of hypothesis testing.

Hypothesis	Path	Path coefficient	$t$ value	$P$ value	Hypothesis supported?
H1	PA $\rightarrow$ AIR	0.237	3.560	0.000	Y
H2	I $\rightarrow$ AIR	0.056	1.045	0.094	N
H3	RL $\rightarrow$ AIR	0.075	2.368	0.003	Y
H4	LftV $\rightarrow$ AIR	0.008	1.981	0.037	Y

## 5. Conclusion

Athlete's injury rehabilitation has always been a hot topic in academic circles. Scholars try to come up with solutions that help athletes recover from injuries from different perspectives. Based on the perspective of the internal environment, this paper constructs a model of factors that affect athletes' injury rehabilitation. In this article, we focus on the four factors of psychological adjustment, imagery, rehabilitation learning, and video reflection. Through questionnaire investigation and empirical analysis, we found that psychological adjustment, rehabilitation learning, and video reflection have a significant positive impact on athletes' injury rehabilitation, while imagery has no significant impact.

Our research provides a reference plan for athletes to adopt effective injury rehabilitation training methods. At the same time, we have also enriched the research literature on injury rehabilitation solutions. We emphasize that athletes need to maintain a good attitude during the injury

repair period. For example, they can make psychological adjustments and release stress by talking with coaches, family members, and communicating. At the same time, the injury recovery period is also a period of learning and reflection. Athletes should seize this time without losing the opportunity to improve their theoretical level through knowledge learning and video analysis.

Our research has undeniable limitations. First, this article only discusses four factors that may affect athletes' injury rehabilitation based on the internal environment perspective. In the future, a variety of factors can be analyzed based on different perspectives. Second, the research sample in this article is 129, so the research results have sample limitations. In the future, when conditions are mature, the sample size can be increased, and relevant factors can be explored. Finally, this article uses an online questionnaire collection method, so it is difficult to completely guarantee the authenticity of the surveyed information. In the future, offline or a combination of online and offline methods can be used to collect questionnaires.

## Data Availability

The data used to support the findings of this study are included within the article.

## Conflicts of Interest

All the authors do not have any possible conflicts of interest.

## References

- [1] R. Hare, L. Evans, and N. Callow, "Imagery use during rehabilitation from injury: a case study of an elite athlete," *The Sport Psychologist*, vol. 22, no. 4, pp. 405–422, 2008.
- [2] D. Forsdyke, A. Smith, M. Jones, and A. Gledhill, "Psychosocial factors associated with outcomes of sports injury rehabilitation in competitive athletes: a mixed studies systematic review," *British Journal of Sports Medicine*, vol. 50, no. 9, pp. 537–544, 2016.
- [3] X. Chen and G. Yuan, "Sports injury rehabilitation intervention algorithm based on visual analysis technology," *Mobile Information Systems*, vol. 2021, Article ID 9993677, 8 pages, 2021.
- [4] C. Sordoni, C. Hall, and L. Forwell, "The use of imagery by athletes during injury rehabilitation," *Journal of Sport Rehabilitation*, vol. 9, no. 4, pp. 329–338, 2000.
- [5] H. Dhillon, S. Dhillon, and M. S. Dhillon, "Current concepts in sports injury rehabilitation," *Indian journal of orthopaedics*, vol. 51, no. 5, pp. 529–536, 2017.
- [6] J. King, C. Roberts, S. Hard, and C. L. Ardern, "Want to improve return to sport outcomes following injury? Empower, engage, provide feedback and be transparent: 4 habits!," *British Journal of Sports Medicine*, vol. 53, no. 9, pp. 526–527, 2019.
- [7] K. D. Peterson and L. C. Evans, "Decision support system for mitigating athletic injuries," *International Journal of Computer Science in Sport*, vol. 18, no. 1, pp. 45–63, 2019.
- [8] M. Case, *Athletic Injuries: Psychological Impacts and Athlete Preparedness to Return to Participation*, Informa Healthcare, 2019.
- [9] S. Abdollahi and R. Sheikhhoseini, "Sport-related injuries in Iranian basketball players: evidence from a retrospective epidemiologic study (2019–20)," *The Physician and Sportsmedicine*, 2021.
- [10] T. Nakamura and H. Tsuchiya, "Association of hardness and social support with posttraumatic growth following athletic injuries," *International Journal of Sport and Health Science*, vol. 18, pp. 28–38, 2020.
- [11] A. N. Marshall, T. C. Valovich McLeod, and K. C. Lam, "Characteristics of injuries occurring during cross-country: a report from the athletic training practice-based research network," *Journal of Athletic Training*, vol. 55, no. 12, pp. 1230–1238, 2020.
- [12] M. Stick-Mueller and M. Tunning, "The ICA Athletic Commission: an early effort to connect athletes to chiropractic," *Chiropractic History*, vol. 41, no. 1, 2021.
- [13] R. Elwood, O. El-Hakeem, Y. Singh, H. Shoman, O. Weiss, and V. Khanduja, "Outcomes and rate of return to play in elite athletes following arthroscopic surgery of the hip," *International Orthopaedics*, pp. 1–11, 2021.
- [14] G. Lei, "Application and research of nano-biomaterials in knee cruciate ligament rehabilitation after sports injury," *International Journal of Nanotechnology*, vol. 18, no. 1/2/3/4, pp. 82–96, 2021.
- [15] S. Su, "Statistical calculation method and analysis of athletes' biorhythm state and sports injury," *Mobile Information Systems*, vol. 2021, 9 pages, 2021.
- [16] A. Joseph, S. Gupta, Y. C. Wang, and K. Schoefer, "Corporate rebranding: an internal perspective," *Journal of Business Research*, vol. 130, pp. 709–723, 2021.
- [17] A. Gledhill and A. Ivarsson, "Believe in your ability to create change: psychosocial factors influencing sports injury rehabilitation adherence," *The Psychology of Sports Injury*, pp. 93–106, 2021.
- [18] S. Pierce, *The Importance of Psychosocial Considerations in the Injury Rehabilitation of Elite Athletes*, A Senior Thesis Submitted in Partial Fulfillment of the Requirements for Graduation, 2021.
- [19] D. Forsdyke, *Seeing Is Believing: the Role of Imagery in Sports Injury Rehabilitation*, Taylor & Francis Group, 2021.
- [20] W. Chu, *Studies on the Effects of Wiring Density on Chip Package Interaction and Design Optimization with Machine Learning [Ph.D. thesis]*, 2021.
- [21] J. Zhang, J. Sun, J. Wang, and X. G. Yue, "Visual object tracking based on residual network and cascaded correlation filters," *Journal of Ambient Intelligence and Humanized Computing*, vol. 30, pp. 1–4, 2020.
- [22] Y. Gu, A. Chen, X. Zhang, C. Fan, K. Li, and J. Shen, "Deep learning based cell classification in imaging flow cytometer," *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 2, pp. 18–27, 2021.
- [23] S. Qi, X. Ning, G. Yang et al., "Review of multi-view 3D object recognition methods based on deep learning," *Displays*, vol. 14, article 102053, 2021.
- [24] M. Gao, W. Cai, and R. Liu, "AGTH-net: attention-based graph convolution-guided third-order hourglass network for sports video classification," *Journal of Healthcare Engineering*, vol. 2021, article 8517161, 10 pages, 2021.
- [25] C. Yan, G. Pang, X. Bai et al., "Beyond triplet loss: person re-identification with fine-grained difference-aware pairwise loss," *IEEE Transactions on Multimedia*, vol. 31, 2021.

- [26] D. K. Ahern and B. A. Lohr, "Psychosocial factors in sports injury rehabilitation," *Clinics in Sports Medicine*, vol. 16, no. 4, pp. 755–768, 1997.
- [27] L. Podlog, J. Heil, and S. Schulte, "Psychosocial factors in sports injury rehabilitation and return to play," *Physical Medicine and Rehabilitation Clinics*, vol. 25, no. 4, pp. 915–930, 2014.
- [28] L. B. Green, "The use of imagery in the rehabilitation of injured athletes," *The Sport Psychologist*, vol. 6, no. 4, pp. 416–428, 1992.
- [29] B. Law, M. Driediger, C. Hall, and L. Forwell, "Imagery use, perceived pain, limb functioning and satisfaction in athletic injury rehabilitation," *New Zealand Journal of Physiotherapy*, vol. 34, no. 1, 2006.
- [30] E. Jespersen, "Peers as resources for learning: a situated learning approach to adapted physical activity in rehabilitation," *Adapted Physical Activity Quarterly*, vol. 25, no. 3, pp. 208–227, 2008.
- [31] A. Eastman and D. G. Chang, "Return to learn: a review of cognitive rest versus rehabilitation after sports concussion," *NeuroRehabilitation*, vol. 37, no. 2, pp. 235–244, 2015.
- [32] J. R. Wang and N. Parameswaran, "Survey of sports video analysis: research issues and applications," in *Proceedings of the Pan-Sydney area workshop on Visual information processing*, pp. 87–90, Sydney, 2004.
- [33] I. E. Allen and C. A. Seaman, "Likert scales and data analyses," *Quality Progress*, vol. 40, no. 7, pp. 64–65, 2007.
- [34] B. Hou, W. Huanfang, and X. Guojie, "Study on the influencing factors of crowdsourcing logistics under sharing economy," *Management Review*, vol. 31, no. 8, p. 219, 2019.
- [35] C. Ringle, D. Da Silva, and D. Bido, "Structural equation modeling with the Smartpls," *Brazilian Journal of Marketing*, vol. 13, no. 2, 2014.
- [36] H. M. T. Thaker, A. Khaliq, A. A. Mand, H. I. Hussain, M. A. B. M. T. Thaker, and A. B. A. Pitchay, "Exploring the drivers of social media marketing in Malaysian Islamic banks: an analysis via smart PLS approach," *Journal of Islamic Marketing*, 2020.
- [37] V. Silaparasetti, G. V. R. Rao, and F. R. Khan, "Structural equation modeling analysis using smart pls to assess the Occupational Health and Safety (OHS) factors on workers' behavior. Structural equation modeling analysis using smart PLS to assess the Occupational Health and Safety (OHS) factors on workers' behavior (July 17, 2017)," *Humanities & Social Science Reviews*, pp. 2395–7654, 2017.
- [38] P. U. Gio and E. Rosmaini, *Belajar Olah Data dengan SPSS, Minitab, R, Microsoft Excel, EVIEWS, LISREL, AMOS, dan SmartPLS*, Universitas Sumatera Utara, Medan, Indonesia, 2016.
- [39] K. K. K. Wong, "Partial least squares structural equation modeling (PLS-SEM) techniques using SmartPLS," *Marketing Bulletin*, vol. 24, no. 1, pp. 1–32, 2013.

## Retraction

# Retracted: Construction and Analysis of Green Investment Risk Evaluation Index System Based on Information Entropy Fuzzy Hierarchical Analysis Model

### Wireless Communications and Mobile Computing

Received 8 August 2023; Accepted 8 August 2023; Published 9 August 2023

Copyright © 2023 Wireless Communications and Mobile Computing. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their

agreement or disagreement to this retraction. We have kept a record of any response received.

### References

- [1] Y. Huang, "Construction and Analysis of Green Investment Risk Evaluation Index System Based on Information Entropy Fuzzy Hierarchical Analysis Model," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 4850321, 13 pages, 2021.

## Research Article

# Construction and Analysis of Green Investment Risk Evaluation Index System Based on Information Entropy Fuzzy Hierarchical Analysis Model

Yuqi Huang 

Shaoxing University Yuanpei College, Shaoxing 312000, China

Correspondence should be addressed to Yuqi Huang; [huangyuqi1212@usx.edu.cn](mailto:huangyuqi1212@usx.edu.cn)

Received 15 July 2021; Revised 1 August 2021; Accepted 9 August 2021; Published 30 August 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Yuqi Huang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper adopts the information entropy fuzzy hierarchical analysis model to conduct an in-depth study and analysis of the green investment risk evaluation system and to process and analyze its indicators. The entropy method-fuzzy hierarchical analysis is used to evaluate the green financial capital operation risk. Firstly, I point out the research trends and shortcomings, define the concepts of fund operation, operation risk, risk evaluation, and green finance, and elaborate the relevant basic theories to lay the theoretical foundation for the later research; secondly, I identify the external and internal fund operation risks by describing the development plan, overall operation, and specific operation plan of green finance; then, according to the existing fund operation risks, I integrate commercial banks' risk. Then, based on the existing fund operation risks, I integrate the risk types of commercial banks, combine the characteristics of green finance business, design the green finance fund operation risk evaluation index system, initially determine the index weights through hierarchical analysis method, introduce the entropy value method to improve the index weights, and use the fuzzy comprehensive evaluation method to obtain the final evaluation results: in the process of green finance fund operation, internal risk is the most important risk, among which credit risk and financial innovation risk have the most significant impact; external risk is the second most important risk, among which economic risk and political risk are the most important. External risks are the second most important risks, among which economic risks and political risks are more obvious. The risks are judged from single-factor aspects, and the scoring values are used to determine the higher possibility of schedule risks caused by human factors, and finally, the schedule risk level of the reservoir project is determined to be medium through a comprehensive evaluation. The reasonableness of the calculation results shows that the entropy theory and the fuzzy comprehensive evaluation method are still applicable to the assessment of green investment risks.

## 1. Introduction

As market competition continues to become more complex, many world-renowned companies have been acquired or gone bankrupt due to failures in risk management. To effectively assist enterprises in risk management, the continuous establishment and improvement of the risk management system have played a certain role in guiding the risk evaluation of enterprises, but due to the different risks faced by each industry, it is difficult to adopt a unified framework or guidelines to the specific requirements of different enterprises, and the establishment of a risk evaluation model suit-

able for industry characteristics can solve this problem to a certain extent and break through the quantitative evaluation of risk [1]. The technical bottleneck of quantitative risk evaluation can be broken. The importance of risk evaluation in academic and practical circles contributes to its importance in case studies and is the fundamental reason for choosing risk evaluation as the research topic in this paper. This paper introduces risk indicators reflecting the characteristics of green finance into the risk evaluation index system of commercial banks' capital operation, establishes a scientific risk evaluation index system of green finance capital operation, and provides a theoretical basis for improving the risk

management system of commercial banks and green finance risk research. As an emerging investment and financing tool, equity crowdfunding has the dual attributes of Internet and finance, and the sources of risk are relatively extensive, and the risk factors are more complex, and there are relatively few academic results that quantitatively study the risk of equity crowdfunding [2]. This paper identified and evaluated the investment risks in equity crowdfunding, classified and controlled the sensitivity of various potential risk indicators, and provided new analysis perspectives, ideas, and methods for the risk management theory of Internet equity crowdfunding. In addition, investors, as important participants in the crowdfunding market, face the greatest risks due to information asymmetry, but there is a lack of research in the literature to evaluate and control the risks of crowdfunding from investors' perspective [3]. Therefore, this paper analyzes what types of risks investors encounter in the process of equity crowdfunding investment, how to quantitatively assess the risk level of equity crowdfunding projects with scientific models, how to rank the risk level of several crowdfunding projects with scientific methods, how to identify which risk indicators have a greater impact on investment returns with scientific models, and how this paper identifies and evaluates the risk level of equity crowdfunding projects. This paper identifies and evaluates the investment risks in equity crowdfunding, classifies and controls the sensitivity of various potential risk indicators, and provides a new analytical perspective, ideas, and methods for the risk management theory of Internet equity crowdfunding.

By combining green financial policies with the development of the financial industry, this paper systematizes the formation process of China's green financial policy system and analyzes the performance and shortcomings of green financial policies in the context of the development of the green financial industry while better understanding the green financial policy system. Commercial banks' business object is capital, and all business activities revolve around the capital. The rational flow and optimal allocation of capital elements are used to improve the efficiency of capital use and achieve maximum economic benefits with the smallest capital cost, thereby ensuring the stable, sustained, and efficient development of commercial banks. This paper considers that commercial banks' capital operation is the general term for commercial banks' activities such as financing, investment, and capital operation. Commercial banks obtain funds through financing, gain income through investing, and ensure the orderly implementation of financing and investment activities through the operation of funds. The purpose of the capital operation of commercial banks is to make effective use of funds to achieve the business objectives of commercial banks, to improve the efficiency of capital use through reasonable flow and optimal allocation of capital elements [4], to achieve maximum economic benefits with minimum capital costs, and to ensure stable, sustainable, and efficient development of commercial banks [5]. The management activities of commercial banks will ultimately affect capital management, and the risks faced by commercial banks will ultimately be reflected in the operational risks of capital. Risk evaluation refers to the process of consider-

ing the probability of occurrence of the risk and the degree of loss based on risk identification and risk estimation, combining the principles of comprehensiveness, objectivity, and importance, compared with the recognized safety indicators, assessing the possibility of occurrence of the risk and the degree of harm, and determining whether to take risk reduction measures. At present, the mainstream risk evaluation method is divided into the subjective assignment evaluation method, which determines the weight through subjective judgment scoring; the objective assignment evaluation method which determines the weight according to the relationship between indicators or variation coefficient. The mainstream risk evaluation method has a wide range of applications, such as avionics systems [6], wireless sensor networks [7], and UAV systems [8].

While traditional finance focuses on economic benefits, green finance emphasizes the harmonious development of economic and social benefits. Green finance starts from protecting resources and the environment, effectively managing pollution problems and promoting sustainable economic development, analyzes and judges whether the requirements of balanced economic and social benefits are met, and takes the degree of protecting the environment and saving resources as an important basis for measuring the effectiveness of operation, to achieve a win-win situation for financial development, environmental improvement, and resource-saving. The elements in the risk evaluation system are interconnected and interact with each other and jointly influence the enterprise risk. In this paper, when designing the risk index system, I consider both internal and external risks. External risks include four aspects of policy risk, legal risk, economic risk, and market risk, which represent the main external risks that may be faced, and internal risks include strategic risk, credit risk, liquidity risk, operational risk, financial innovation risk, and reputation risk, which represent the main internal risks that may be faced, not only focusing on risk evaluation as an organic whole but also considering the interaction between risk evaluation and other activities of the company, which better explains the idea of system theory.

## 2. Current Status of Research

In terms of investment risk research, Shaktawat and Vadhera argue that equity crowdfunding provides investors with more options, but micro and small enterprises are at the beginning stage of their business and have a higher percentage of failed crowdfunding investment projects [9]. Liu et al. argue that equity crowdfunding investment is risky, but the actual participation of the investing public has limited investment ability [10]. Bhowmik et al. argue that nonprofessional investors like to blindly follow hot projects and need to pay attention to the potential herding effect [11]. Peng et al. point out that the traditional equity-based high-risk investment nonprofessional investors can use the services of professional investment institutions, while the public involved in crowdfunding cannot scientifically assess the actual value of the project, and is at a disadvantage in the current valuation game, facing the risk of overvaluation of

the underlying [12]. For example, there are no clear and fixed boundaries for such rubrics as investment risky, moderate risk, and small risk. Internal risks include strategic risk, credit risk, liquidity risk, operational risk, financial innovation risk, and reputation risk, which represent the main internal risks that may be faced, and not only focus on considering risk evaluation as an organic whole.

Therefore, the investment risk evaluation of equity crowdfunding is a multiattribute decision problem with a high degree of fuzziness. Therefore, this paper adopts the fuzzy comprehensive evaluation method to evaluate the risk level of equity crowdfunding projects, based on interval intuitionistic fuzzy set theory to rank several projects' crowdfunding investment risks [13]. At present, the research on fuzzy comprehensive evaluation method is relatively mature and widely used in many fields. Considering the different categories of risks in Internet finance in practical situations, Cao Lingyan combined hierarchical analysis with fuzzy mathematics to assess the overall risk level of current Internet finance in China. Wang et al. proposed an index system for evaluating the innovation performance of scientific research institutions in Guizhou province using a fuzzy comprehensive evaluation method and designed an evaluation model based on this index system [14]. He et al. completed the study of beer taste harmonization by selecting several comprehensive evaluation indexes that can describe the taste of beer and introducing the fuzzy comprehensive evaluation method [15]. One of the current research hotspots in the field of fuzzy evaluation is the interval intuitionistic fuzzy multiattribute decision method [16]. The classical fuzzy set theory suffers from numerous limitations in practice due to the small amount of embedded information expressed. In this regard, the intuitionistic fuzzy set theory is proposed, and three concepts of affiliation, nonaffiliation, and hesitation are defined [17]. To solve the problem that information such as affiliation degree and nonaffiliation degree sometimes cannot be portrayed by quasireal values, the intuitionistic fuzzy set is further extended, and the theoretical system of interval fuzzy set is proposed. The advantage of this set is that both affiliation and nonaffiliation degrees are interval values, which can portray fuzziness more flexibly, so the related theory is also often used to solve multiattribute evaluation problems such as investment evaluation [18].

The theoretical and applied research on risk evaluation for the three major activities of financing, investment, and operation of enterprises is mature, and the evaluation of different risks in different industries from various perspectives is realized by combining various mathematical models. The evaluation of risk is comprehensive and focused, while the evaluation of specific risks is targeted and practical. Foreign scholars are more mature in both qualitative and quantitative evaluation of risks, and the research in recent years has focused on introducing new methods to improve the original risk evaluation models to enhance quantitative accuracy and thus adapt to the characteristics of industries and risks, but the improved models are too complex and reduce the practicality. Based on foreign risk management systems, Chinese scholars are keen to establish evaluation index systems for risks in various industries from different

perspectives and to verify the reliability of the systems through empirical analyses using different methods, which has realized the transformation from qualitative analysis to a combination of qualitative and quantitative analysis, but the evaluation results are highly subjective due to the limitations of research methods. Then, construct a weighted standardized evaluation matrix according to TOPSIS. The model determines the maximum value and minimum value of each index of the ideal solution and negative ideal solution according to the maximum value, calculates the Euclidean distance, and then, uses the calculated distance to observe the relative sticking progress, and is based on the relative closeness. Due to the special characteristics of commercial banks' capital operations, it is difficult to accurately divide financing activities, investment activities, and operation activities, so this paper divides risks for collation, determines the main structure of this paper's risk evaluation index system, introduces entropy value method-fuzzy hierarchical analysis method to process relevant data, reduces the subjectivity of fuzzy hierarchical analysis method in determining index weights, and improves the research results. The reliability of the research results is improved.

### 3. Information Entropy Fuzzy Hierarchical Analysis Model for Green Investment Risk Evaluation Index System Construction

*3.1. Information Entropy Fuzzy Hierarchical Analysis Model Design.* Mathematicians were the first to publish the concept of information entropy [19, 20], laying a solid foundation for quantitative information analysis, providing a detailed explanation of entropy theory, constructing the basic model, and writing it into its mathematical theory of propagation [21]. The concept of thermal entropy in physics is an index used to measure the disorder of molecular states. Shannon adopted the concept of information entropy to perfectly explain the uncertain information source, and the amount of information entropy is used to determine the amount of information. The specific definition of entropy is as follows: let a discrete random variable  $X = \{x_1, x_2, \dots, x_n\}$ . The information source included given appears with  $p_i = p(x_i)$  probability and  $\sum_{i=1}^n p_i = 2$ , then the information entropy of  $X$  can be expressed by the formula.

$$H(x) = \sum_{i=1}^n p_i \ln \frac{1}{p_i}. \quad (1)$$

By calculating the information entropy, the size of the information can be known, and the original information can also be reflected objectively. There is an inverse proportional relationship between information entropy and the amount of information; the greater the information entropy, the smaller the information of the index, and the smaller the role produced by the comprehensive evaluation, and vice versa, which shows the important significance of information entropy in the evaluation. Similarly, the smaller the information entropy, the greater the amount of information provided by the index, and the greater its contribution to the

comprehensive evaluation results. Therefore, the weight of indexes in the comprehensive evaluation results is larger.

In this paper, the risk evaluation matrix of green building developers is constructed by selecting the secondary indicators of each evaluation object from four aspects: political risk, economic risk, social risk, and technical risk according to the actual cases, conducting a comprehensive scientific analysis of the detected raw data, further processing the model criteria, carrying out entropy values and entropy weights for different indicators, and determining the degree of influence of each indicator on the risk evaluation. Then, a weighted standardized evaluation matrix is constructed according to TOPSIS [22]. Everyone's boundary definition of fuzzy things is not the same, which shows that it is necessary for us to maintain a certain degree of subjectivity when solving fuzzy problems. Even for the treatment of objective things, ambiguity is applicable. The model determines the maximum and minimum values of each indicator of the ideal solution and negative ideal solution evaluation object based on the maximum value, by the calculation of the Euclidean distance, and then uses the calculated distance to observe the relative fit progress and based on the relative fit. The higher the ranking, the lower the risk, and vice versa, the higher the risk.

The study object is analyzed for its relevant indicators, and data information is derived for the original matrix data  $M$  with the following formula.

$$M_{(ij)m \times n} = \begin{pmatrix} p_{11} & p_{21} & \cdots & p_{m1} \\ p_{12} & p_{22} & \cdots & p_{m2} \\ \cdots & \cdots & \cdots & \cdots \\ p_{1n} & p_{2n} & \cdots & p_{mn} \end{pmatrix}. \quad (2)$$

The authorized matrix  $P_{(ij)m \times n}$  is obtained after doing uniform standard processing on the matrix of the study subject, and the formula is as follows.

$$P_{(ij)m \times n} = \begin{pmatrix} p_{11} & p_{21} & \cdots & p_{m1} \\ p_{12} & p_{22} & \cdots & p_{m2} \\ \cdots & \cdots & \cdots & \cdots \\ p_{1n} & p_{2n} & \cdots & p_{mn} \end{pmatrix}. \quad (3)$$

The entropy value of each index has been determined by applying the entropy weighting formula to calculate the respective entropy value.

$$e_j = -k \sum_{i=1}^n p_i \ln \frac{1}{p_i}. \quad (4)$$

The weights calculated by entropy weighting are combined with TOPSIS to form a weighted normalization matrix  $V$ , which is calculated for each indicator with different levels

of importance in the risk assessment.

$$V = (X_i P_{ij})^{mn}, \quad (5)$$

$$(V_{ij})_{mn} = \begin{pmatrix} X_1 p_{11} & X_1 p_{21} & \cdots & X_1 p_{m1} \\ X_2 p_{12} & X_2 p_{22} & \cdots & X_2 p_{m2} \\ \cdots & \cdots & \cdots & \cdots \\ X_m p_{1n} & X_m p_{2n} & \cdots & X_m p_{mn} \end{pmatrix}. \quad (6)$$

When the factor set  $U$  contains a lot of data, the relative membership degree weight coefficient will be smaller when the weight vector sum is 1. The incomplete consideration of the index information of each evaluation object in the fuzzy comprehensive evaluation process will also affect the evaluation result distinction. The risk evaluation index system should involve the internal and external risks that may be faced in the operation of green finance funds, but there are many risks faced, and it is difficult to highlight the focus of prevention and control if all the risks are included, so the risk evaluation index system should be designed by selecting representative indicators according to the current situation of risks, while avoiding the mutual influence of each indicator, taking into account the focus and comprehensiveness, so that the risk evaluation index system will have practicality and guidance. The system is shown in Figure 1.

Intelligent technology [23–25] is one of the main tools to achieve risk evaluation and risk management. Quantitative analysis can visually measure the risk level, but the quality of data required for quantitative analysis directly affects the credibility of risk evaluation, and it is difficult to achieve quantitative evaluation if there is insufficient data [26]. Qualitative analysis, on the other hand, is an empirical judgment of indicators, which can be used as a reference for quantitative analysis, but care should be taken to avoid the influence of its subjectivity as much as possible. Corporate loans have grown significantly. The balance of green project loans accounted for a slight decline in company's loan balance, but remained at about 10%. The environmental impact assessment rate of loan projects was 100%. Therefore, the process of green financial capital operation risk evaluation should combine the advantages of quantitative and qualitative evaluation to reduce the influence brought by insufficient data and strong subjectivity.

A fuzzy concept means that the boundary of this concept has uncertainty, and it can also be said that its boundary is unclear and vague. Ambiguity is an objective property of things, and the uncertainty of the state of the event itself, i.e., it is difficult to determine whether the object of study meets the requirements of a certain concept. It should be noted that when people perceive vagueness, it is subjective, which means that everyone does not have the same boundary limits for vague things, which means that it is also necessary for us to maintain some subjectivity when solving vague problems. Even for the treatment of objective things, vagueness is applicable.

Fuzzy mathematics is a theoretical approach to deal with fuzzy phenomena, and its research is about the uncertainty



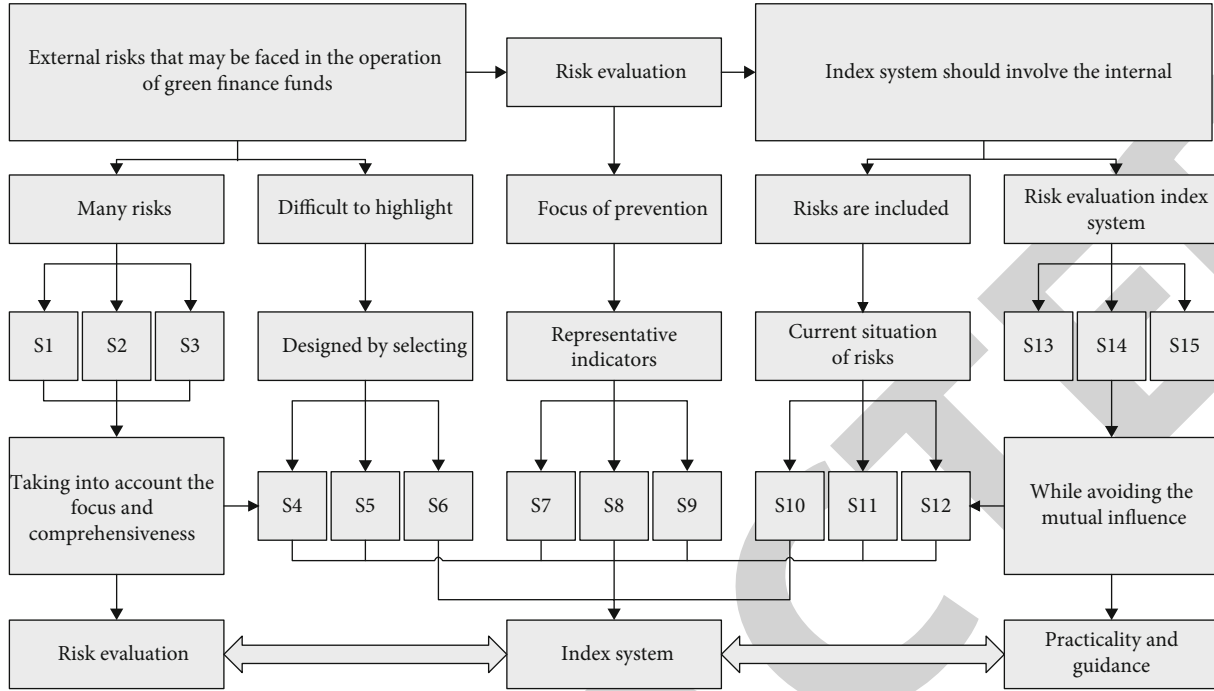


FIGURE 1: Information entropy fuzzy hierarchical analysis model.

of things. It replaces classical sets with fuzzy sets and thus extends the concepts in classical mathematics so that mathematical methods can be used not only to solve exact problems but also to play an important role in fuzzy areas. Since fuzzy concepts can be described utilizing fuzzy sets, fuzzy mathematical methods can also be used to solve problems when applying these concepts for judgment, evaluation, reasoning, decision-making, and control, such as fuzzy cluster analysis, fuzzy pattern recognition, fuzzy comprehensive judgment, fuzzy decision-making, fuzzy prediction, fuzzy control, and fuzzy information processing. The fuzzy comprehensive evaluation method used in this paper is based on fuzzy mathematics, and through the quantitative analysis of fuzzy risk factors, the risk level of the evaluated object is comprehensively evaluated from the perspective of multiple indicators.

The basic principle of the fuzzy comprehensive evaluation method is to first determine the set of factors and evaluation set of the evaluated object, and then use a suitable method to calculate the weights of each factor separately, and establish a fuzzy relationship matrix by the determined affiliation degree, and then, synthesize and normalize the fuzzy relationship matrix and the weight vector by the fuzzy operator to get the comprehensive evaluation results. The overall framework of the fuzzy comprehensive evaluation method is shown in Figure 2.

For multiple evaluated objects, the weighted average method can sort them based on their rank positions. If the elements in the rubric set  $V$  are not quantified, each  $v_j$  needs to be quantified first when the weighted average method is used. After obtaining the fuzzy comprehensive evaluation results, to make full use of the information contained in  $B$ , the evaluation results and the rubric set can be considered

together to make the evaluation results more realistic, i.e., the corresponding evaluation grade can be determined by calculating the comprehensive score of each evaluation object.

$$F = B^2 V^T. \tag{7}$$

When there are more influencing factors involved in the judging object, it is difficult to determine the allocation of weights reasonably, which requires the use of multilevel fuzzy comprehensive judging. The results of the first-level fuzzy evaluation are composed into a new evaluation set for the second-level or multilevel fuzzy comprehensive evaluation, and the final evaluation results are obtained by superimposing the evaluation calculations at all levels. The fuzzy comprehensive evaluation method is used for all the evaluation objects, and the evaluation objects are ranked by the comprehensive score through the data processing of the corresponding evaluation indexes. The fuzzy comprehensive evaluation is based on the status of the affiliation of multiple indicators to the evaluated thing, divides the change interval of the evaluated object, and reflects the fuzziness of the evaluation criteria and influencing factors. The fuzzy comprehensive evaluation result is ultimately a vector, which contains rich information and not only can reflect the fuzzy condition of the thing itself more accurately but also can be further processed to get reference information. In dealing with fuzzy evaluation objects, the fuzzy comprehensive evaluation method can make more scientific and reasonable, close to the actual quantitative evaluation, especially in the subjective evaluation, the evaluation process of fuzzy comprehensive evaluation method can make full use

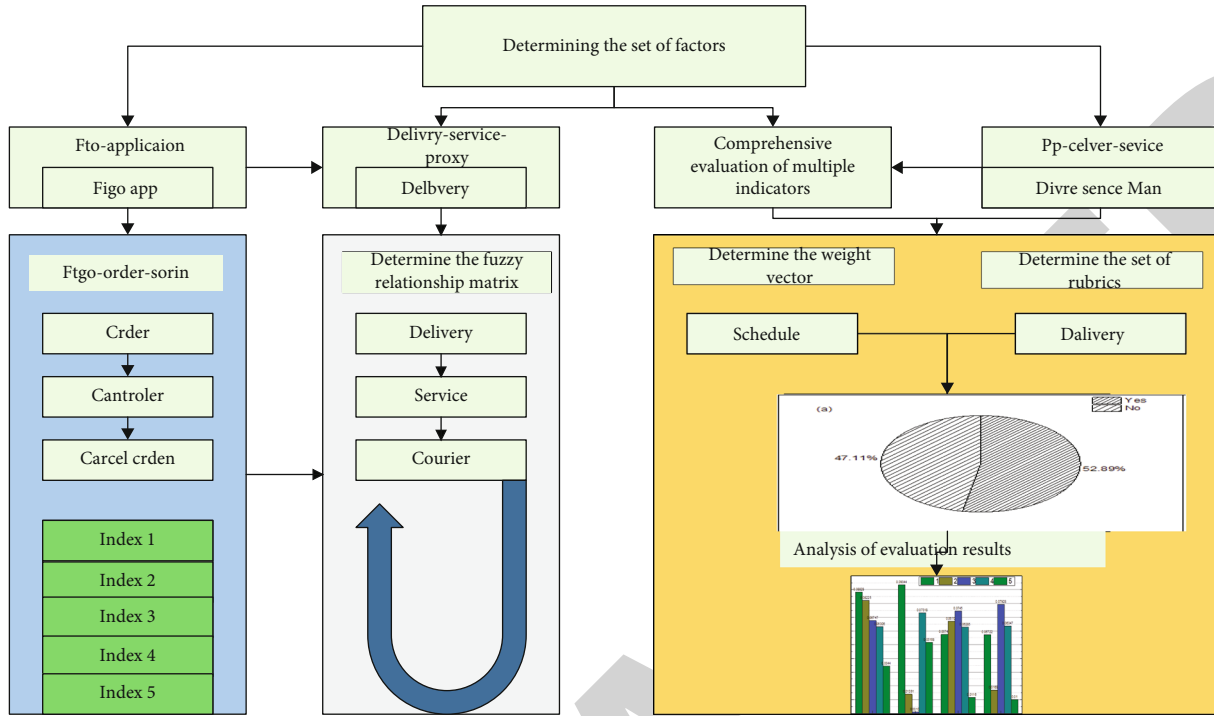


FIGURE 2: The overall framework of the fuzzy integrated judging method.

of the index data and has wide applicability. First, the relevant background, ideas, methods, indicator design, goals, and other content of this research are explained to the experts, and the experts are required to evaluate the relevant green finance policies based on the evaluation indicators and calculate them in accordance with the corresponding preset indicators. Evaluation results of various policies.

There are shortcomings in the application of the fuzzy comprehensive evaluation method; first, it does not solve the problem of duplication of information caused by the relevant evaluation indexes, and it is more subjective in the determination of the index weight vector. In addition, when the factor set  $U$  contains more data, the relative affiliation coefficient will be small when the weight vector sum is 1. The incomplete consideration of the information of each evaluation object in the fuzzy comprehensive evaluation process will also affect the differentiation of the evaluation results. The fuzzy comprehensive evaluation method provides a new mathematical tool to solve the problem of multi-indicator evaluation and decision-making in a fuzzy environment, and the reasonable determination of the weights of each indicator is an important step to get reliable evaluation results. The weight is a quantitative value that compares and weighs the relative importance of each factor in the overall evaluation. It can be done through expert questionnaires and other simple and easy ways. In addition, the meaning of the evaluation index must be accurate, the data sorting must be standardized, and there must be corresponding quantitative methods available. Indicator weights can be calculated by dividing multiple levels, and to reflect the importance of each indicator in the established evaluation index system, different weight coefficients must be

assigned to each indicator. For the same set of indicator values in the comprehensive evaluation process, different weights can lead to completely different or even opposite evaluation conclusions.

**3.2. Green Investment Risk Evaluation Index System Construction.** In the green finance business, I implement all green finance requirements in all aspects of due diligence, review, and approval, and postaward inspection, and establish a long-term mechanism for green finance [27]. I focus on supporting the development of industries such as the environmental protection industry and clean energy generation and include the environmental protection industry in the active support category, insisting on a preferential selection of regions, active entry, control of maturity, and locking of loan repayment. Large-scale lending in projects such as green transportation projects, building energy-saving, and green building and industrial energy-saving and water-saving, and environmental protection with good development prospect at present and the loan balance of green finance projects from 2015 to 2017 is outlined as shown in Figure 3.

Against the backdrop of intensifying competition in the banking industry, the balance of green project loans grew year by year with positive YoY growth rates. Due to rapid business development, company's loans grew significantly, and the balance of green project loans accounted for a slight decrease in company's loan balance, but remained at around 10%, with a loan project EIA rate of 100%. While formulating credit policies to support the development of the environmental protection industry, I continue to innovate the green financial service system and take the lead in developing green financial services such as CDM financial advisory

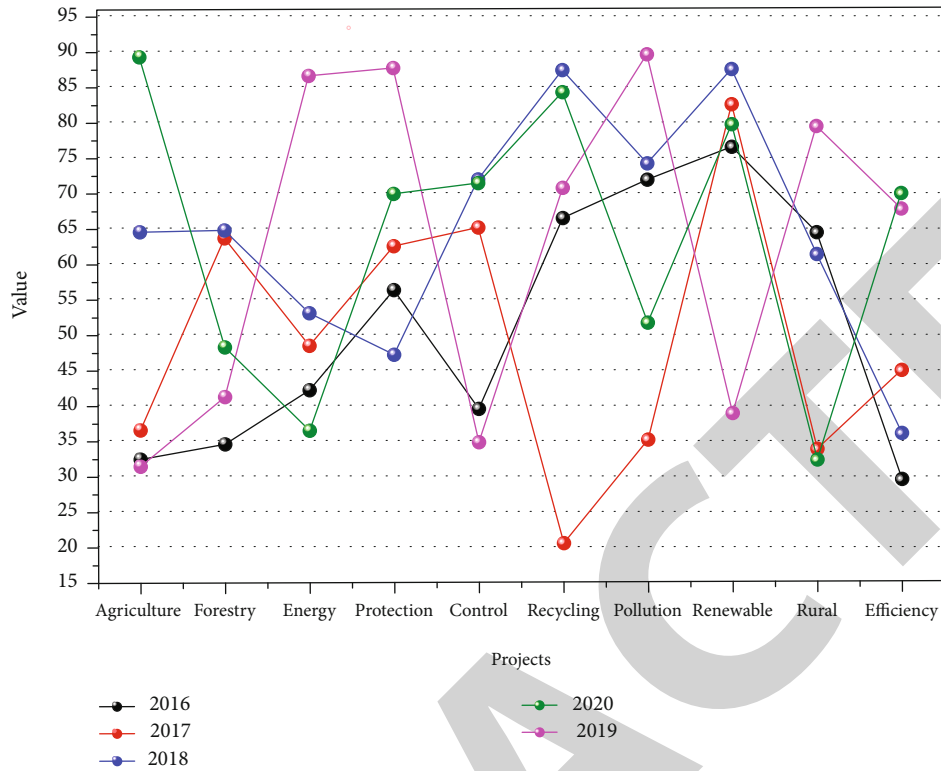


FIGURE 3: Green project balance.

business, international carbon finance collateral business, contract energy management factoring financing, future income rights pledge financing, building energy efficiency financing products, offshore escrow of certified emissions, carbon bonds, and domestic certified voluntary emission reduction pledge financing loans, and other green financial services for specific. To carry out effective risk response strategies to the risks existing in the project development process, formulate practical and reasonable risk response strategies. I adopt different financial services and mortgages for specific projects to help environmentally friendly enterprises obtain financial support through various financing channels, reduce financing costs, and at the same time diversify operational risks and reduce the probability of losses.

Loans in high energy-consuming and high-polluting industries are mainly concentrated in chemical raw materials and chemical products manufacturing and ferrous metal smelting and rolling processing industries, which together account for a three-year average of 70.68% of loans in high energy-consuming and high-polluting industries, and the overall loan stock in high energy-consuming and high-polluting industries has not decreased significantly, but with the significant increase in company’s loan balance, the loan stock in high energy-consuming and high-polluting industries account for a significant declined. Loans in overcapacity industries were mainly concentrated in the iron and steel industry and the cement industry, which together accounted for 78.99% of the loans in overcapacity industries on average over three years. Due to the significant increase in total loans, the proportion of loan stock in the “two high

and one surplus” industries decreased by 1.13 percentage points [28]. It is necessary to strengthen the training of itself and management personnel. At the same time, it is necessary to train and assess the staff engaged in green building development and management at each stage. At the same time, some managers with relevant green building experience can be hired from outside to manage the project. The environmental risk control of loans to chemical raw materials and chemical products manufacturing, ferrous metal smelting and rolling processing industry, iron and steel industry, and cement industry is crucial to reduce the loan balance of “two high and one surplus” industries and promote industrial transformation.

Government’s financial investment in energy conservation and pollution prevention and control scale and tax incentives continue to rise, but due to the lack of integration of state financial investment, mostly for short-term financial support rather than long-term investment planning, the supervision of financial investment is difficult to ensure. Many environmental projects are not profitable enough on their own, the project construction early to meet environmental standards to obtain many state subsidies and tax incentives, but high maintenance costs later, the sustainability and stability of state financial support, and tax incentives directly affect the risk of capital investment. Tolerance of industries destroys the environment seriously but contributes to tax revenue. The supervision and management of green finance still exist imperfect systems, unclear division of authority and responsibility, policy enforceability that cannot be guaranteed, and insufficient motivation of

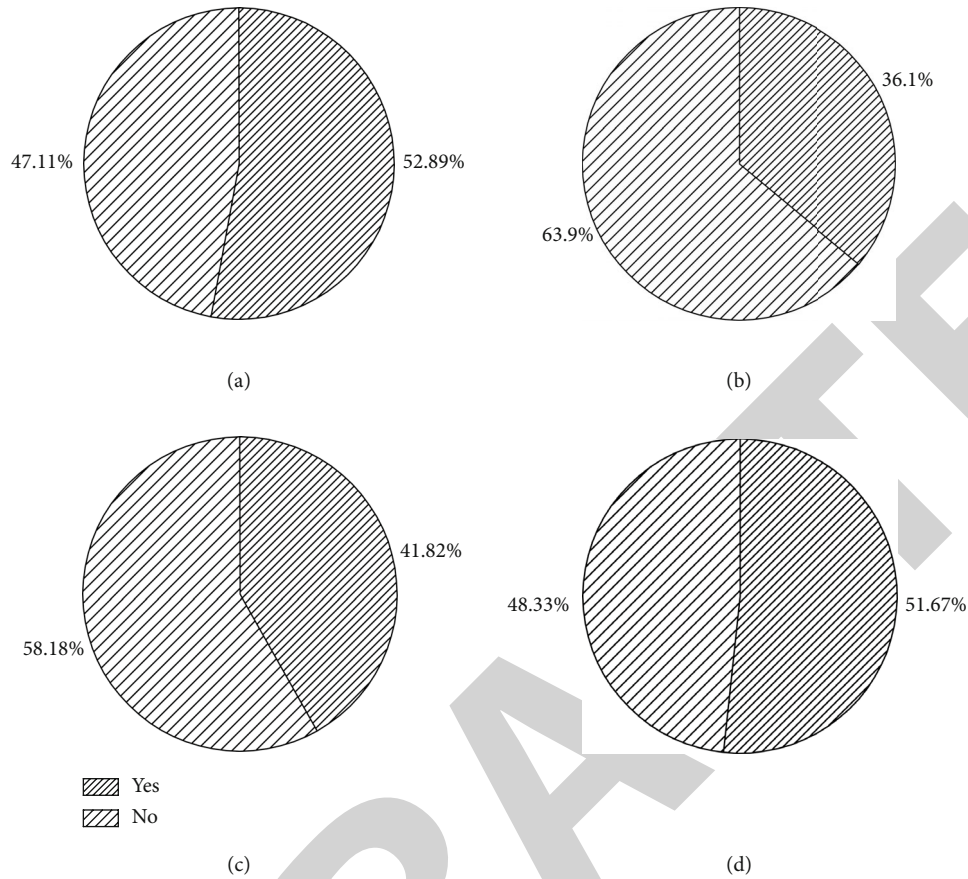


FIGURE 4: Number of green investments.

personnel to implement policies, which makes it difficult for many favorable policies of the central government to be put into practice for environmental protection. The development of polluting industries not strictly limited will further weaken the environment. The profitability of the environmental protection industry, which in turn affects the risk of capital recovery, is shown in Figure 4.

The trend of the global real economy slippage is obvious, whether to achieve a turnaround is still difficult to determine, the U.S. economic growth is under great pressure, the U.S.-China trade war is still ongoing, and the intensification of the debt of many countries is the hidden risks in the international economic development. The economy is currently in a critical period of “major adjustment,” economic growth too fast decline. As a member of the financial industry, which is affected by the macroeconomic situation, the international and domestic macroeconomic situation is not optimistic, which directly leads to the reduction of business scale and project capital recovery difficulties.

The design of evaluation indicators is a key part of the evaluation process, and a scientific indicator system is developed to examine whether the indicators can adequately reflect the quality of green finance policies so that the basic characteristics of the evaluation objects can be accurately reflected. Firstly, the background, idea, method, indicator design, and objective of this study are explained to the

experts, and they are asked to evaluate the relevant green financial policies based on the evaluation indicators and calculate them according to the corresponding preset indicators, to obtain the evaluation results of various policies.

The more indicators are not the better, but the selected influencing factors should be representative, play an important role in the evaluation process, and reflect the characteristics of the evaluation object [29]. The establishment of the indicator system is also simplified as much as possible, either by qualitative analysis of the interrelationships among the indicators or by using quantitative methods to select representative indicators to achieve the optimization of the indicator system. This can reduce the time and workload of evaluation, and the distribution of the weights of each factor is visible. The factors at the same level of the indicator system should be independent of each other, not duplicated as much as possible, and there is no causal relationship between them, and the system is hierarchical. The evaluation index system needs to be established according to the hierarchical development of the comprehensive evaluation process, and satisfactory evaluation results can be obtained. The established index system should be able to be applied to the evaluation of water conservancy projects, so the selection of indicators should be in line with objective reality and have a reliable source of data, through expert questionnaires and other easy-to-use ways. In addition, the meaning of

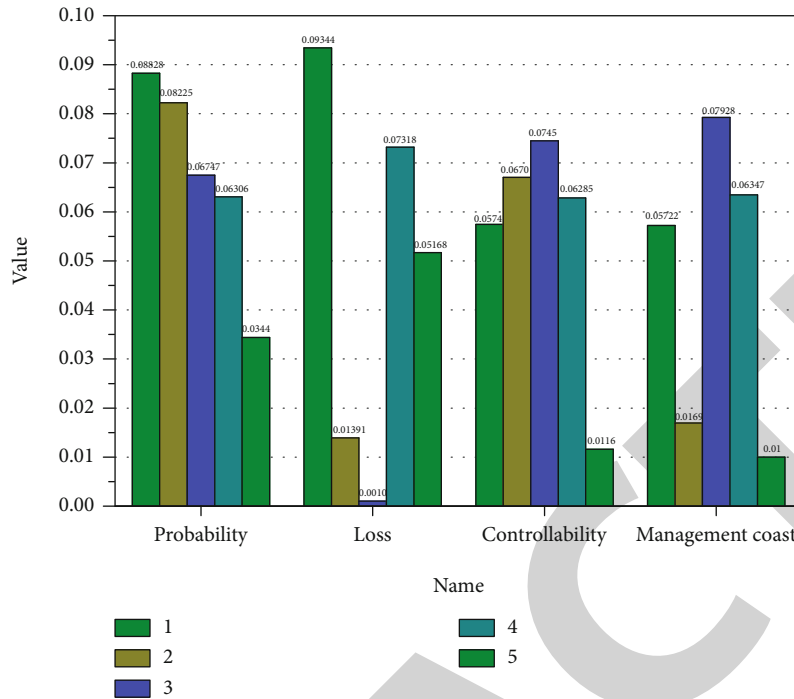


FIGURE 5: Risk weighting.

evaluation indicators should be accurate, the data should be standardized, and in addition, the corresponding quantification methods can be used.

#### 4. Analysis of Results

**4.1. Model Performance Results.** The raw data were obtained by interviewing experts and scoring risk factors, and the data were first normalized. A total of 30 questionnaires were distributed through electronic questionnaires to project managers and interviews with experts in this area of green building, and all responses were received, of which 28 were valid. To complete the development of green building projects, the risks of the projects need to be controlled effectively. To make an effective risk response strategy for the risks existing in the project development process, a practical and reasonable risk response strategy is formulated. However, for the developer of the green building project, the resources, time, and human resources needed in the project process are limited; to the effect of limited resources, the problems that will be encountered in the whole development process, and the risks faced, these factors that appear should be analyzed specifically, analyze the size of the impact of the risk on the development process, and at the same time, combined with the entropy right and TOPSIS in the project. At the same time, the top three risk factors in the project are selected as the key assessment objects, and effective prevention and management methods are given to the risk factors to better deal with the risks. The top three risk factors were selected from four aspects and summarized in a table (as shown in Figure 5).

At present, the visibility of green buildings has also achieved a certain status in the construction market, but

for the public, there is also no better understanding of the green building. One of the key issues in the green building field is how to get the public to properly recognize and accept it. At present, as the green building industry in China is still in its initial stage, the development is not mature enough, and the market is not perfect. The public subconsciously believes that green building is high-grade building, and there are some misunderstandings about the cost of green building. To improve public acceptance, the public needs to have a correct understanding of green building; first, the W project development department of company B can obtain public's concerns and questions about green building through the social questionnaire. According to the ranking of political risk factors, the relative closeness of imperfect laws and regulations of green building is 0.8152.

Then, I can use the huge advantage of the Internet to design an APP to release the data related to development and operation on the Internet so that the public can have a better understanding of the whole process of green building and at the same time cooperate with relevant media to promote the connotation of green building and green building-related products, and establish our brand to gain public recognition. In the green building industry, since developers do not have enough management experience, if they want to engage in the development of green building projects and succeed, they need to strengthen the training of themselves and their managers, and at the same time, they need to train and assess the staff engaged in the development and management of green building at each stage, and at the same time, they can hire some managers from outside who have relevant green building experience to manage the projects (as shown in Figure 6).

The actual green building project as a case, according to the risk list, is in line with the actual situation; using the

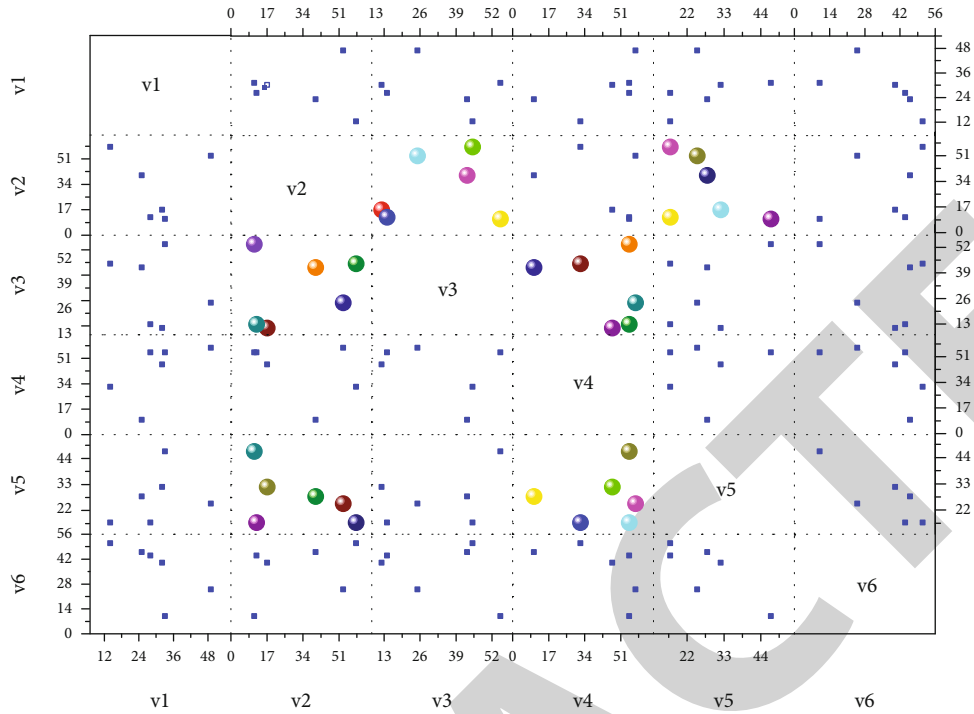


FIGURE 6: Mutation detection results.

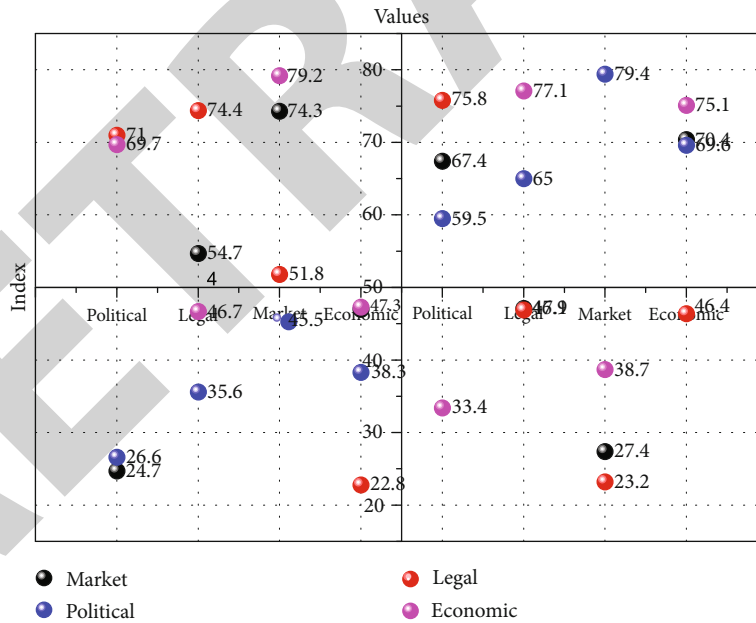


FIGURE 7: Indicator weights.

expert interview method, the risk index system is in line with the risk list of the project, according to the indicators of the risk list issued questionnaires to evaluate the 18 risk indicators, political risk as an example, and according to the entropy weight method and TOPSIS combined with the evaluation, the ranking of political risk factors. The relative proximity of imperfect laws and regulations of green building is 0.8152, the proximity of approval procedure and effi-

ciency of administrative departments to the project implementation is 0.7798, and the proximity of corrupt behavior and bureaucracy of the government is 0.5326. The top three risk coping strategies are given, considering the policies of city A and the actual operation mode of city B. On the premise of obtaining the evaluation result vector Bui of each level index, the preliminary judgment on the risk level of the construction period is made from different

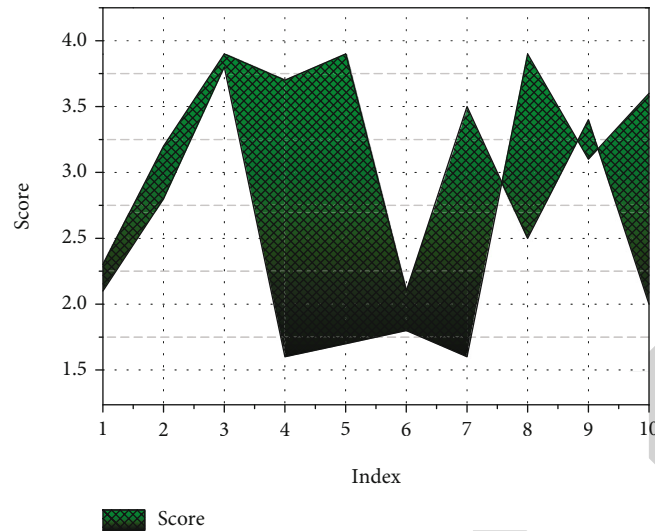


FIGURE 8: Risk indicator scoring chart.

perspectives according to the principle of maximum subordination.

**4.2. Indicator Evaluation Results.** The results of the experts' scoring of the secondary indicators in the external risk were aggregated, and the weights were calculated as shown in Figure 7.

As a business operating in money, the macroeconomic growth rate has a clear impact on the number of customers and the scale of business of a single customer, while the growth prospects and profit margins of customers affect the most critical repayment factor—the level of cash flow. When the macroeconomic trend is good and stable, they usually face lower external risks; thus, they will lower the customer entry threshold and increase the scale of financial support to earn spreads on a large scale, benefiting from both benefits and risks; once the economic growth rate drops steeply, the large-scale financial support originally provided with lowered standards in a good economic situation will face serious credit losses. Thus, economic growth rate and stability have a huge impact on the risk of green financial fund operation.

The proportion of “two high and one leftover” industries in the industrial structure affects the exit speed. Due to the profit-oriented, the huge market scale of “two high and one leftover” industries in China makes it difficult to withdraw from the market quickly. Although the head office will issue a certain stock exit target and incremental scale limit every year, subordinate branches will usually meet the assessment requirements by posting lines to meet the target or manipulating statements, etc. Compared with the huge base, the exit speed is slow. A large proportion of the environmental industry in the industrial structure has a positive effect on green finance business, but the development within the environmental industry is seriously unbalanced, and the immaturity of some green projects leads to small market size and poor project profitability, which in turn affects the enthusiasm for financial support.

However, with huge risk exposure and insignificant profitability, it is difficult to guarantee the initiative to develop green finance. Therefore, the establishment of a sound risk mechanism for green finance has a significant impact on improving returns, reducing risks, and thus improving the initiative to develop green finance; as the risk characteristics of green finance projects are significantly different from those of traditional business, the government is gradually building a regulatory system adapted to the characteristics of green finance. It is bound to face serious regulatory requirements.

The final scores of each level of indicators can also be represented more visually by creating bar charts (as shown in Figure 8). Due to the strong policy relevance and high technical content of the green finance business, it is difficult for nonspecialized departments and personnel to control the risk and nodes of green projects in the whole process of operation. Although the separation of incompatible positions and authorization and approval system is set in the business process, branches at all levels may lower the entry and risk review standards of the green finance business and neglect the implementation of the exit policy of the “two high and one leftover” industries out of the consideration of customer relationship and performance, making the green finance process a mere formality. Although the whole process of business informatization has been realized, key information and approval still require manual intervention, and business and management personnel at all levels may help customers “greenwash” or weaken pollution data to obtain financial services out of their interests or under pressure from superiors.

Any problems and risks faced in the operation of green finance funds will affect reputation risk to varying degrees, and the public stakeholders are numerous, so reputation risk in the operation of green finance funds cannot be ignored. Green financial brand influence, social responsibility, and green financial publicity have a positive impact on reputation risk. The comprehensive service strength in the field

of green finance determines brand influence, while the development of green finance is one aspect of its social responsibility, and green financial publicity affects the ability to promote green financial services. Strengthening all other aspects of risk management is also a part of reputation risk management. It is difficult for nonprofessional departments and personnel to control the risks and nodes of the whole process of green projects in the operation process. Although the separation of incompatible duties and authorization and approval systems are set up in the business process, branches at all levels may be out of consideration for maintaining customer relationships and performance.

The internal risk of green financial operation risk is 4.26, which is between very likely and very likely, which is consistent with the late start of green financial business, insufficient development enthusiasm, and imperfect national policy and regulatory system in China. The overall evaluation of external risk in green financial fund operation risk is 4.05, which is between very likely and very likely, which is consistent with the current situation that green financial development is strongly policy-oriented, and commercial banks are influenced by economic and market risks. To cope with liquidity risk, it is necessary to enrich the financing channels of the green finance business and increase the proportion of specialized green finance financing channels in the investment. To cope with strategic risks, it is necessary to formulate strategic plans in line with the objective environment, supervise the implementation of the strategies, and ensure the consistency of the strategies of the head office and branches. To deal with operational risks, it is necessary to set up a dedicated institution for green finance, allocate professional staff, and improve employees' risk awareness. To deal with reputation risk, we should actively assume social responsibility and promote the concept of green finance through various media channels, thereby improving social influence.

## 5. Conclusion

Based on the current situation of green finance business development, I comprehensively analyzed the main risks it may face in the process of fund operation, firstly introduced AHP-fuzzy comprehensive assessment method to rate the risks of crowdfunding investment projects, then focused on improving the traditional intuitionistic fuzzy ranking model, and proposed an interval intuitionistic fuzzy ranking method considering experts' hesitation preference to realize the risk ranking of several crowdfunding projects. Specifically, it is divided into risk factors such as funding method, project access criteria, development speed, industry configuration, and dedicated institutional staff. A green finance fund operation risk evaluation index system is constructed. This paper conducts a comprehensive analysis through four external risk perspectives and six internal risk perspectives and finally comes up with the possibility of each risk occurring in the process of green financial fund operation and gives corresponding suggestions. The weights of secondary and primary indicators are improved by the entropy value method, which reduces the subjectivity of the fuzzy hierar-

chical analysis model in determining the weights of indicators and improves the reliability of the research results. The validation of the constructed evaluation system by only one case is inevitably insufficient. In the future, the evaluation of green financial capital operation risk of domestic commercial banks should take the commercial banks developing green financial business as the research object to ensure the universality of the research results.

## Data Availability

The data used to support the findings of this study are included within the article.

## Conflicts of Interest

The author does not have any possible conflicts of interest.

## References

- [1] X. J. Li, "Research on investment risk influence factors of prefabricated building projects," *Journal of Civil Engineering and Management*, vol. 26, no. 7, pp. 599–613, 2020.
- [2] X. Wang, H. Zhao, and K. Bi, "The measurement of green finance index and the development forecast of green finance in China," *Environmental and Ecological Statistics*, vol. 28, no. 2, pp. 263–285, 2021.
- [3] J. Lu, S. Zhang, J. Wu, and Y. Wei, "COPRAS method for multiple attribute group decision making under picture fuzzy environment and their application to green supplier selection," *Technological and Economic Development of Economy*, vol. 27, no. 2, pp. 369–385, 2021.
- [4] J. Zhang, X. Jin, J. Sun, J. Wang, and K. Li, "Dual model learning combined with multiple feature selection for accurate visual tracking," *IEEE Access*, vol. 7, pp. 43956–43969, 2019.
- [5] X. Liu, "Green production benefit evaluation model of trade products based on principal component analysis," *International Journal of Product Development*, vol. 24, no. 2/3, pp. 217–234, 2020.
- [6] Z. Luo, "Application and development of electronic computers in aero engine design and manufacture," *ASP Transactions on Computers*, vol. 1, no. 1, pp. 6–11, 2021.
- [7] Y. Li and J. Cao, "WSN node optimal deployment algorithm based on adaptive binary particle swarm optimization," *ASP Transactions on Internet of Things*, vol. 1, no. 1, pp. 1–8, 2021.
- [8] J. Chen, C. du, Y. Zhang, P. Han, and W. Wei, "A clustering-based coverage path planning method for autonomous heterogeneous UAVs," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–11, 2021, In press.
- [9] A. Shaktawat and S. Vadhera, "Risk management of hydro-power projects for sustainable development: a review," *Environment, Development and Sustainability*, vol. 23, no. 1, pp. 45–76, 2021.
- [10] X. Liu, G. Tian, A. M. Fathollahi-Fard, and M. Mojtahedi, "Evaluation of ship's green degree using a novel hybrid approach combining group fuzzy entropy and cloud technique for the order of preference by similarity to the ideal solution theory," *Clean Technologies and Environmental Policy*, vol. 22, no. 2, pp. 493–512, 2020.
- [11] C. Bhowmik, M. A. Kaviani, A. Ray, and L. Ocampo, "An integrated entropy-TOPSIS methodology for evaluating green



## Research Article

# Optimization Study of Multidimensional Big Data Matrix Model in Enterprise Performance Evaluation System

Honglin Fu 

*Hong Kong Baptist University, Kowloon Tong, Kowloon, Hong Kong*

Correspondence should be addressed to Honglin Fu; fuhonglin0903@163.com

Received 15 July 2021; Revised 1 August 2021; Accepted 10 August 2021; Published 27 August 2021

Academic Editor: Shan Zhong

Copyright © 2021 Honglin Fu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper uses a multidimensional big data matrix model to optimize the analysis and conduct a systematic construction of the enterprise performance evaluation system. The adoption of new research methods and perspectives to promote the study of the use of performance information is of great significance to achieve the effectiveness, science, and sustainability of corporate performance management. To solve the problem of objectivity and scientificity of performance information use, this part attempts to analyze performance information use from the perspective of the multidimensional big data matrix, focusing on the techniques and methods in the process of promoting performance information use from the multidimensional big data matrix and tries to construct a system model of enterprise performance information use from two dimensions: the use of performance information sources and the use of performance information results. Based on multiple theoretical hypotheses, a theoretical and empirical basis is provided for the division of demand dimensions of enterprise performance evaluation system. Through social capital theory, three dimensions of network social capital, cognitive social capital, and structural social capital are hypothesized, and the logistic regression method is applied for empirical study. The results show that these three dimensions have significant effects on the knowledge demand of enterprise performance evaluation systems. It is verified that the multidimensional big data matrix can enhance the quality of performance information sources and improve the objectivity of performance information. In the performance information source use dimension, the analysis verified that the collection and preprocessing technology of big data can realize the automation, real-time, and diversification of information collection and preprocessing, and enhance the objectivity of performance information. Big data helps to improve the quality and effectiveness of performance information results use. In the dimension of using performance information results, the distributed computing and analysis processing technology of big data can assist the decision support system, and the use of information can be shifted from micromanagement to decision support, to realize the scientific use of performance information and improve the quality of enterprise management decisions.

## 1. Introduction

The importance of research on corporate performance management and related topics has received continued academic attention. Corporate performance management, which draws on the methods of business administration, is a new administrative model developed from the Western New Public Management. Since the rise of research, the theory of corporate performance management has developed rapidly, especially in the areas of performance index systems, performance assessment subjects, and performance assess-

ment tools. Fruitful research results have been achieved [1]. The academic research results of corporate performance management have been on a growing trend in the last decade or so. Corporate performance management has been one of the important issues in public administration research and a hotspot and key area in the discipline of public management, and public management academics have devoted greater attention and power to research on the topic of corporate performance management [2]. With the popularity of the Internet, artificial intelligence, and 5G technology, big data has entered all occupations. It has not only

brought great convenience to individuals' lives but also provided new directions for enterprises to think about. Small and medium-sized enterprises are the main body of the structure of the market economy, with a large number, small scale, limited capital, weak risk resistance, and other characteristics [3]. If SMEs want to be invincible in the fierce competition of the market economy, they must adapt to the changes of the times, use big data technology to assist the development of enterprises, and improve their competitiveness. Corporate performance management has always been one of the important issues of public administration research, and it is also a hot and key field of public management disciplines. Public management academic circles have devoted greater attention and strength to the research on corporate performance management topics. As an important part of enterprise management, human resource management has a direct impact on whether the strategic planning of enterprises can be realized. As the core function of HR management, performance management determines the level of HR management of an enterprise [4]. The development of big data technology brings greater opportunities and challenges to HRM. How to use big data to collect, organize and analyze macro- and microeconomic information to help realize company strategy, promote management level, and make the enterprise invincible in the fierce market competition has become a new issue.

However, the management of enterprises in the context of big data is also facing many new issues. One of the more prominent issues is the performance management of employees and teams. First, the staff and job settings and work processes of the appraisal team in the context of big data are very different from the previous ones. Their main work tasks are real estate data mining, database management, and updating, data platform development and construction, network maintenance, and promotion and after-sales of data products. The difference in work tasks inevitably makes the performance appraisal indexes of employees very different. Traditional performance appraisal indexes focus more on employees' assessment experience and skills, while the use of data assessment systems should focus more on personnel's ability of data analysis and software development. Second, the association between jobs in the context of big data is stronger and requires a higher degree of teamwork. Work tasks include real estate data mining, database management and update, data platform development and construction, network maintenance, and data product promotion and after-sales. Therefore, the traditional performance appraisal system focuses more on the assessment of individual performance, while the performance appraisal of the team should be more important in the context of big data [4]. Again, big data requires the assessment system to have nationwide massive data resources, which inevitably brings about the expansion and development of enterprise organizational structure. The expanding organizational structure brings great difficulty to the performance management of personnel. The headquarters of the enterprise and its subordinate branches as well as the project teams are scattered among many cities in the country, so it is difficult to use a unified performance index assessment system for per-

formance management, and the quality of personnel and job positions is very different among different teams. Therefore, the performance appraisal in the context of big data should consider both the correlation between the work of each team and the differences in the economic development of each region [5]. Therefore, the performance appraisal system in the context of big data should be a complex system considering various factors.

To sum up, the traditional performance appraisal system can no longer adapt to the current industry in the context of big data. In the context of big data, the strategic objectives of enterprises have changed a lot, shifting from the original traditional assessment method to the direction of data platform assessment. Therefore, the performance management system of the enterprise should also reflect the strategic objectives of the enterprise. In this way, the value of talents can be better assessed and staff can be motivated to work creatively. Through a scientific and systematic performance management system, managers can understand the overall team's work status and work quality, find problems in time, solve problems, and help managers make correct decisions, which can further improve the company's management level, while a scientific performance appraisal system can also provide a reference basis for staff salary increases and job promotions and help managers discover the company's outstanding talents. The scientific and systematic performance management system can clarify the development direction of the company and the competition between teams, which can greatly improve the work efficiency, and let the employees see that they will get the corresponding reasonable compensation for their hard work, so that they can stimulate their work enthusiasm, continuously improve their workability, and contribute to the further development of the company.

## 2. Current Status of Research

The calculation of weights among various performance indicators has been a difficult problem, and there has always been a lack of uniform calculation standards; therefore, attempts have been made to solve the calculation of weights using the development of corresponding software [6]. On the enterprise side, Intel has proposed a 360-degree performance evaluation appraisal method. Its evaluation dimensions are diversified to ensure a more comprehensive assessment, and it is mainly used to appraise personnel above the middle level. Key performance indicators (KPI) are also widely used at this stage. This method is to divide the goals of the enterprise according to different levels and classes and to find out the key factors or key events, which affect the development of the enterprise, to evaluate the performance level of employees, and through the scientific application of the KPI, the method can effectively improve the management level of the enterprise. The key performance indicator appraisal method is characterized by being relatively simple and easy to operate, so it is more widely used in SMEs [7]. In summary, the balanced scorecard method, the key performance indicator method, the 360-degree performance appraisal method, and the goal management method are the commonly used performance appraisal

methods in foreign companies [8]. These four methods have their advantages and shortcomings, as well as their scope of application. There are more in-depth researches in the theory of performance appraisal management, and some large enterprises have made outstanding explorations in the practice of performance appraisal management [9]. The main problem is that the corporate culture in different regions is different, the assessment methods used, the assessment system established and the assessment indexes also differ, and a unified performance assessment management system has not been established. At present, enterprises often use the key performance indicator method, 360-degree performance appraisal method, etc. The full combination of theory and practice and mutual promotion make this part of the research get continuous and in-depth development [10]. And in recent years, foreign countries have begun to pay attention to performance evaluation research in the organizational context. Researchers in performance appraisal emphasize that performance appraisal is seen as a complex process that includes social context, employee feelings, and cognition. Some experts and scholars have also begun to focus on the study of team performance and have achieved some results. Performance appraisal should consider not only the relevance of the work of each team but also the differences in the economic development of each region. Therefore, the performance appraisal system in the context of big data should be a complex system that considers various factors.

Liu et al. emphasize that firms form a large amount of unique, hard-to-replicate tacit knowledge based on practices, skills, and other resources and transform this knowledge into firm competitiveness through effective accumulation and learning within the organization [11]. Based on this, Zhou et al. argue that dynamic capabilities are processes by which firms use resources, and he points out that dynamic capabilities can not only support firms to adjust their strategies to match the changing environment, but it can even create new markets and improve business performance [12]. In a constantly changing environment, it is difficult to bring lasting benefits to a company with unchanging core competencies. Grant et al. explored the mechanism of corporate knowledge. Xiao et al. argue from the perspective of organizational knowledge evolution that enterprise capability is ultimately the accumulation of empirical knowledge of the enterprise, which is reflected in the business processes of the company's business [13]. The theory of enterprise knowledge is an integration of the resource base theory, core competency theory, and dynamic competency theory. Core competence is a heterogeneous resource of enterprises, while knowledge is the basis of the core competence of enterprises, and the difference of competitive advantages of enterprises originates from the different mechanisms of knowledge creation and application [14]. Enterprises should not only consider enhancing the utility of current knowledge but also seek new knowledge acquisition, and constantly updating knowledge is the key to maintain competitive advantage.

Among the theoretical studies, some scholars mainly explore the realization path of innovative enterprise performance management models utilizing big data thinking and

machine learning technologies [15–17]. Perkhofer believes that the characteristics of the era of big data and public sector performance management have a fit and build a simple model of public sector performance management thinking, suggesting scientific use of data mining techniques to build a database system of public sector performance management information [18]. Through the analysis of massive performance data, deeper performance management issues are discovered. It focuses on exploring the influencing factors of applying big data technology [19] in the public sector, proposes a key approach to big data application, presents a theoretical framework to explain the motivation, capability, and performance of big data application in the public sector, and investigates the impact path of big data application on public service performance. It also proposes that future research should construct and test the theoretical framework of the causes, processes, and performance of big data applications in the public sector and make recommendations to promote big data applications in the public sector.

### 3. Multidimensional Big Data Matrix Model in Enterprise Performance Evaluation System Optimization Analysis

*3.1. Optimal Design of Multidimensional Big Data Matrix Model.* Multidimensional data fusion is an automated information processing approach that brings together multiple sources of information to assist decision-makers in making decisions by transforming the collected information into representational values that can be processed [20, 21]. The process of data fusion involves many aspects such as data collection, processing, detection, combination, and evaluation. Through data processing, multisource data is fused, so as to accurately identify the state of the event, the safety of the environment, and the participants' identity and other information. The process of data fusion involves numerous aspects of data collection, processing, detection, combination, and evaluation of data from multiple sources to accurately identify information such as the state of an event, the level of safety of the environment, and the identity of the participants. In sorting out the definition of multidimensional data fusion, it can be found that multidimensional data fusion is the process of organizing data from multiple sources, identifying the corresponding knowledge, and removing redundant and useless information, to obtain target state information and rank the importance of things that need to be processed, thus helping users to make decisions. The theory of multidimensional data fusion was first applied to the military field to deal with target tracking and identification, system warning, and situational assessment in military information.

Multidimensional data fusion has also undergone a long period of development and is also characterized as interdisciplinary and cross-disciplinary, a comprehensive discipline involving knowledge and techniques from various fields. Therefore, scholars do not have the same definition of multidimensional data fusion, and the accepted definition for information fusion is that proposed by the Federation of

Directors of the Tri-Service Organization Laboratory. The fusion of multidimensional data can be performed using Bayesian estimation. When fusing, the data should be as independent as possible. By dividing independently, Bayesian estimation can be used to evaluate the system for decision-making. Scholars have different definitions of multidimensional data fusion. For the definition of information fusion, the universally recognized definition is proposed by the United States Tri-Service Organization Laboratory Council.

Suppose that the decisions that can be made in the system are  $A_1, A_2, \dots, A_m$ , and the observed outcome is denoted as  $B$ . This allows the emergent prior knowledge to be used so that the probability of deciding  $p(A_i)$  and the outcome  $p(B|A_i)$  that occurs under a certain decision can be obtained. The probability of an outcome  $p(B|A_i)$  under the probability of a decision can be transformed into a posteriori probability  $p(A_i|B)$  in the Bayesian formulation. This allows the Bayesian formula to be represented using the conditional formula.

$$p(A_i|B \cup C) = \frac{p(B \cup C|A_i)}{\sum_{j=1}^m p(B \cap C|A_j)}, \quad (1)$$

where  $B$  is an observation and  $C$  is observation under another source. For the decision condition  $A_i$ , the probability of occurrence of observation  $C$ ,  $p(B \cup C|A_i)$ , such that the probability of simultaneous occurrence of  $B$  and  $C$  needs to be calculated. However, this calculation is difficult, so the formula can be further modified. First assume that  $A$ ,  $B$ , and  $C$  are independent of each other.

$$p(B \cup C|A_i) = p(B|A_i)p(C|A_i). \quad (2)$$

Then, the equation is rewritten as equation (3).

$$p(C|A_i \cup B) = \frac{p(B \cup C|A_i)p(C|A_i)}{\sum_{j=1}^m p(B \cap C|A_j)p(B|A_j)}. \quad (3)$$

Rough set theory can effectively deal with some noisy data or provide solutions to problems such as incomplete data and inaccurate classification. In the rough set theory, the processing method for knowledge is more comprehensive, and the knowledge granularity can be divided, to ensure the accuracy of knowledge, which is also a knowledge theory domain method. Through rough set theory, it can provide a theoretical basis for the treatment of some uncertainty problems [22]. The coarse and fine-grained knowledge is also the criterion for determining whether knowledge can be accurately divided and whether it can accurately describe the concept of a certain field. Rough set is used as the dividing basis, and knowledge is defined on this basis, to describe the concept of the domain more accurately. The advantage of the rough set theory is that it does not require a priori information, can effectively classify knowledge with known information, and can approximate a large amount of repetitive knowledge, so that it is easier to find law in this knowledge. This feature makes rough set theory widely used in the fields of data mining, expert systems, and pattern recogni-

tion, and it can effectively analyze the hidden laws or rules in knowledge without adding any information of data.

By these ideas makes rough sets can be better used in the field of data mining. When performing data mining, it is possible to uncover the fine-grained nature of knowledge and uncover associations between data. Knowledge can be partitioned using rough sets, and this partitioning is often based on a theoretical domain. The coarse and fine granularity of knowledge is also the criterion that determines whether knowledge can be accurately classified and whether it can accurately describe the concepts of a domain. Using rough sets as the basis for division, knowledge is defined on this basis to describe the concepts of the domain more accurately, as shown in Figure 1.

Cross-source multidimensional government data sharing is an extension of the traditional concept of government data, and its data-sharing model architecture is also significantly different from the traditional model architecture, mainly including three extensions: the data of business systems are different, including the data of many departments such as public prosecution, law enforcement, government affairs, and industry and commerce administration. When constructing the system, some of the data come from the self-built systems of the units at this level, while some originate from the national and provincial vertical systems, in addition to the interface between the self-built system and the data sharing and exchange platform, the technology of data pipeline is also used to realize the interface between the national vertical system and the data-sharing platform. Part of the data comes from the self-built system of the unit at the same level, and part comes from the national and provincial vertical systems. Therefore, when building the system, in addition to the connection between the self-built system and the data sharing and exchange platform, the data pipeline technology is used to realize the national vertical system, interfacing with the data-sharing platform. The diversity of data structure includes structured, semistructured, and unstructured data, such as structured data of business processing in the governmental system, semistructured data of system logs, and unstructured data of each governmental monitoring system.

$$BEL(A) = \sum m(B)^2. \quad (4)$$

The virtual knowledge community contains a large amount of user behavior data and personalized demand data. These data constitute the multidimensional data of the virtual knowledge community. Using these data for fusion can alleviate the data sparsity and cold-start problems in traditional recommendation systems, and therefore, hybrid recommendation with fused multidimensional data is an important means for a virtual knowledge community to solve user information overload. When dealing with multidimensional data, it often faces situations such as missing data or unknown data, so it is necessary to predict the filling of missing data or unknown data based on known data. Clustering methods can be used to cluster the data and predict the relationship between missing data using existing

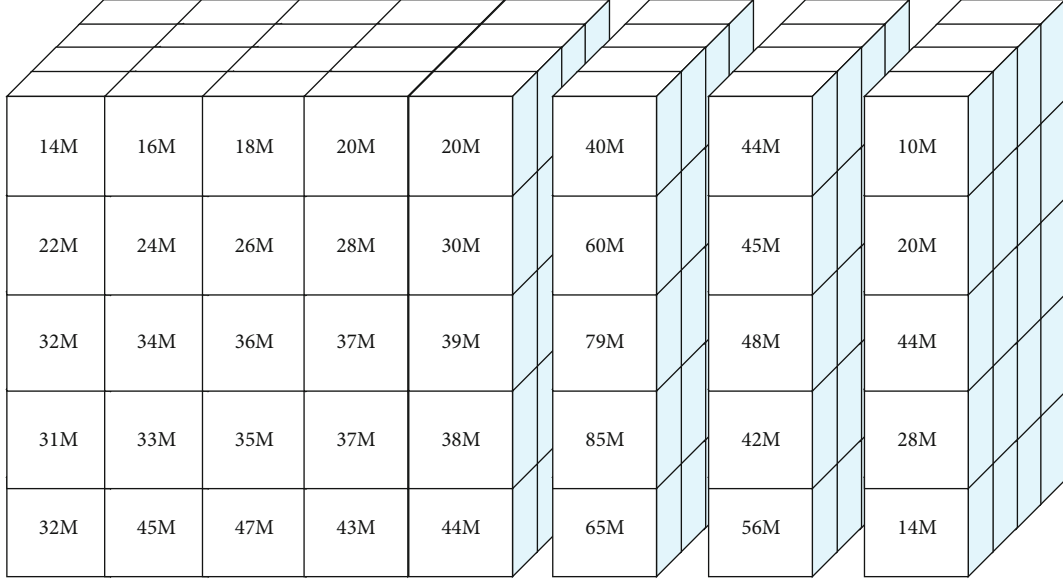


FIGURE 1: Multidimensional big data matrix model.

relationships, thus considering the correlation between data, and the process of calculation is easier to understand. Missing data can also be filled using a method based on the expected maximum. You can also use methods based on the expected maximum to fill in missing data. This is a feature that uses nonnormal distribution and uses a kernel function to calculate the data distribution density. This is a feature that uses nonnormal distributions and calculates the density of the data distribution using kernel functions. The classifier in Bayes and the expected maximum can predict the data to be filled. There are also studies on missing data filling methods based on self-associating neural networks, which can obtain data with higher accuracy.

The cross-source multidimensional government data-sharing influence factors are  $n$ , and each factor is evaluated by  $m$  levels, then the set of influence factors to be classified. Where each  $u$  is a set of factor evaluation vectors, representing the set of influence degree measures of the  $k$ th factor. The study object matrix can be obtained, as shown in equation (5).

$$X_{n \times m} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix}. \quad (5)$$

To reduce the problem of “swallowing small numbers by large numbers” when comparing data of different orders of magnitude, the original data need to be normalized to the interval  $[0,1]$  by data normalization. The  $j$ th column of the matrix  $X$  is normalized, see (6).

$$x'_{ij} = \frac{x_{ij} + \max \{x_{ij}\}}{\max \{x_{ij}\} + \min \{x_{ij}\}}. \quad (6)$$

The reference series and the evaluation object are standardized by the data and the correlation coefficient of the corresponding index between the  $j$ -rated evaluation index of the  $i$ th factor and the reference series in equation (7).

$$\xi_{ij} = \frac{\max \min \{x_{ij}\}}{\max \{x_{ij}\} + \min \{x_{ij}\}}. \quad (7)$$

To calculate the correlation degree, the influence degree of different indicators in the research object needs to be considered comprehensively, so the indicator weights are introduced, as shown in Figure 2, which is the weight of the influence level of each influence factor, and the correlation coefficient of all indicators of each factor can be combined into one correlation degree, and the correlation degree can be obtained after considering the weights.

$$r_i = \sum_{j=1}^m p_j \xi_i(j). \quad (8)$$

Performance information is characterized by the properties of information itself but also has some special characteristics of performance management. The characteristics of performance information include purpose, accuracy, timeliness, formality, consistency, and completeness. The purpose of performance information is that performance information is not blind and random information but has a specific purpose. Accuracy of performance information means that it can accurately fit the purpose of performance management and meet the needs of performance management. The earlier performance information is obtained, the greater the effect; conversely, the later the time, the weaker the function of the information. The formality of performance information is the value of performance information to realize evaluation and judgment. Timeliness can also be understood as timeliness. Performance information reflects the status of

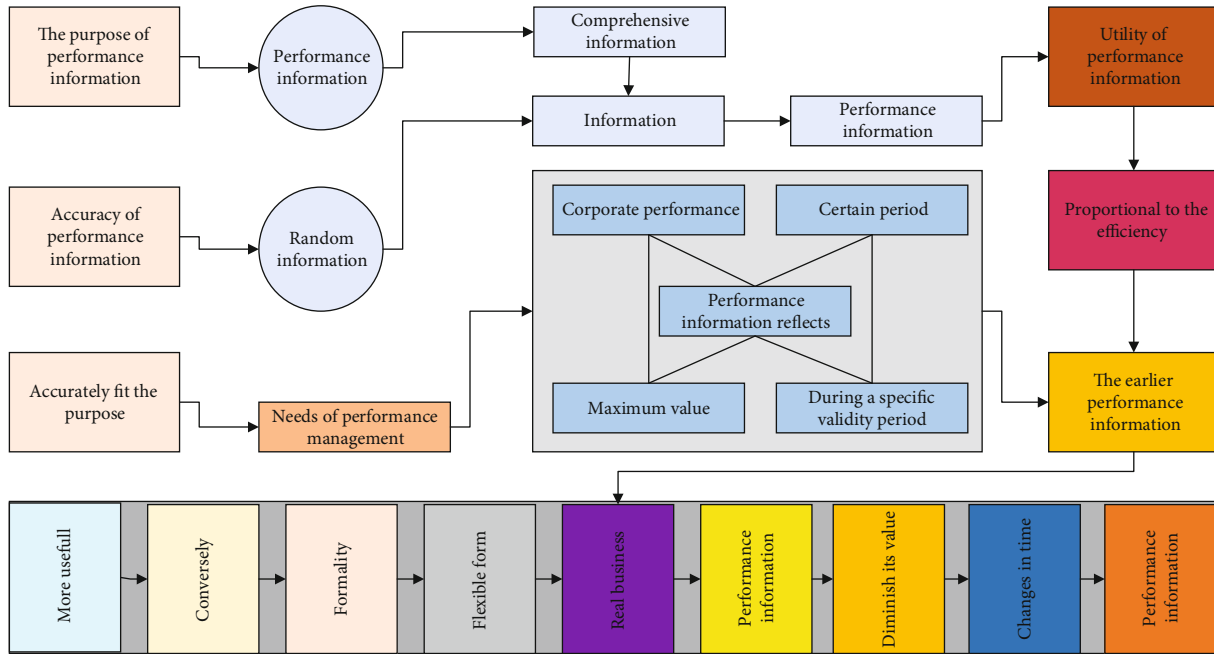


FIGURE 2: Design framework.

corporate performance within a certain period. It has maximum value only during a specific validity period. Therefore, its value is closely related to time, and the utility of performance information is directly proportional to the efficiency of its supply. The earlier performance information is obtained, the more useful it is; conversely, the later it is obtained, the less useful it will be. The formality of performance information is that to achieve the value of evaluation and judgment, performance information needs to be in a clear and flexible form to show the real business management situation in all aspects. Consistency is that performance information does not diminish its value because of changes in time and place. Of course, more importantly, performance information must be complete. The performance of business management requires complete and comprehensive information, so the information cannot be left out; otherwise, it will bring a negative impact on the use of performance information results.

*3.2. Experimental Design of Enterprise Performance Evaluation System.* The goal of performance management based on big data technology is to improve the accuracy and credibility of performance management using big data technology, make counseling more targeted through timely feedback, and ultimately achieve the purpose of mobilizing employees' work motivation and creativity, improving individual performance, and enhancing the overall performance level of the enterprise. Unlike the traditional performance management model, the data source channels of HR performance management based on big data technology are diverse [23]. The source of data information no longer relies on the original data only by manual entry but directly extracts the performance management information of a certain time point automatically through the system. The form

of data information is richer and more diversified, covering information such as graphics, reports, figures, tables, and sounds in addition to conventional text carriers, which expand the dimensionality of data sources and are a fusion of structured data and unstructured data and cannot be concluded using statistical analysis only, so it is necessary to organize, analyze, and store this information through big data technology to complete the informatization process. To be used by managers in general big data is stored on an open cloud platform, which provides a data source for collecting data and sharing data and provides support for the openness of performance management data. The openness is reflected from two aspects: on the one hand, it is open to the users of performance data, providing help to those who need to use the data for decision analysis within the enterprise, reducing the cost of acquiring data for the enterprise, and helping to improve the accuracy of their decisions; on the other hand, it is open to the uploaders, enabling the platform to continuously collect more data and cover a wider range of data, as shown in Figure 3.

The sharing of performance management data can realize data transparency and reduce the cost of acquisition that occurs due to information asymmetry. Data information reaches employees and managers directly, and employees can check their performance results at any time and learn about the performance results of others, so that employees can know themselves and their opponents, which can play the role of self-checking and supervision, and find the gap with others in time and motivate each other to learn and improve work motivation. Performance management based on big data can realize real-time performance feedback and automation. As shown in Figure 3, the performance management based on big data technology can chase the treadle on aj method in real-time when the fruit exceeds the preset

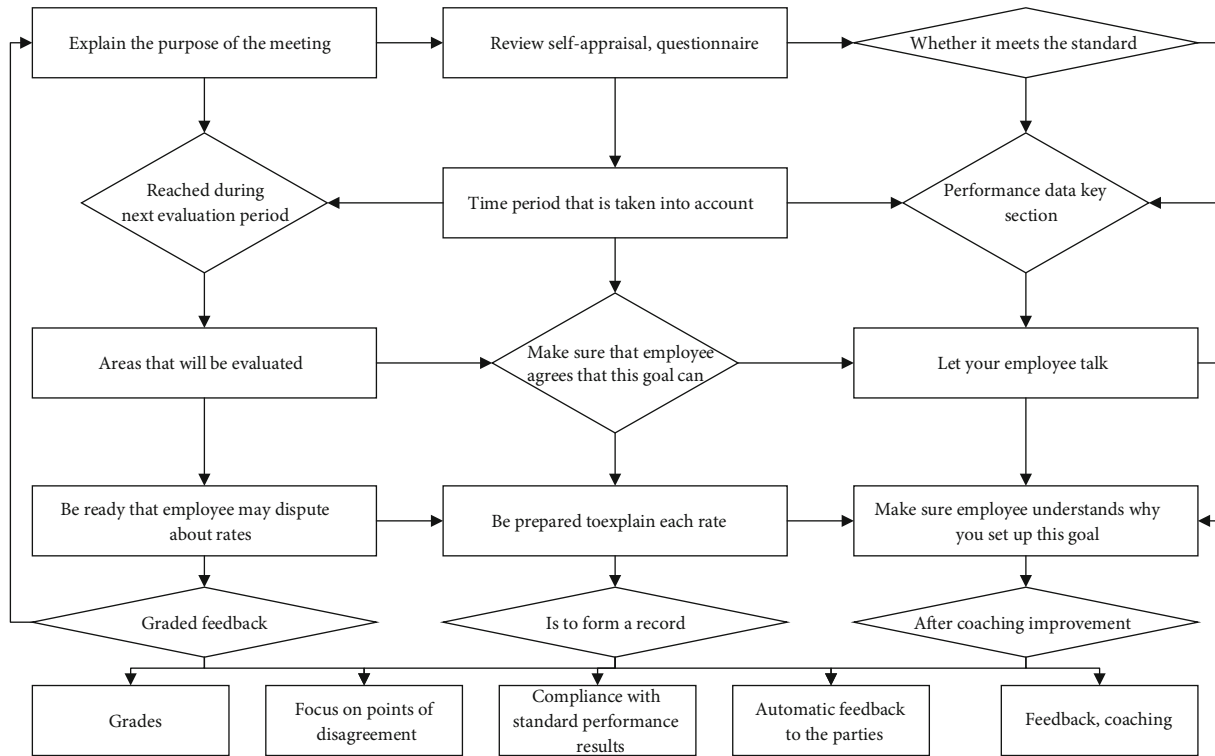


FIGURE 3: Performance feedback process.

value; the system will automatically issue the results for comparison and analysis through human-computer interaction [24, 25]. If the assessment results do not reach the pre-corrected instructions, the system will send the results to the person concerned for guidance and feedback; if the assessment results do not reach the preset indicators, the system will send the results to the supervisor at the same time, so that the supervisor can understand the current situation of employees' work and conduct targeted counseling on time [26].

For the performance appraisal and feedback process, all data are made by database for data custody, collecting more complete information and more accurate appraisal results. The performance appraisal management system based on big data can effectively help employees conduct self-examination, correct bias, and improve work efficiency. Performance management based on big data technology realizes real-time tracking, data collection, comparison, analysis, and feedback so that the level of human resources performance management is greatly improved.

At present, many local enterprises still lack a scientific reward and punishment system, and the incentive validity of performance information use is not high. In terms of incentive direction, there are more positive incentives than negative incentives in the use of performance information for corporate civil servants. In many cases, the results of performance evaluation are generally satisfactory. Among the incentive approaches, the positive incentive of job promotion is significantly more than the other incentive approaches, which has a stimulating effect on civil servants' officialism. Data information forms are more abundant and

diversified. In addition to conventional text carriers, it also covers graphics, reports, numbers, tables, sounds, and other information. These data expand the dimensions of data sources and are the fusion of structured data and unstructured data. The unbalanced and unscientific setting of incentive methods is often unable to properly guide the attitudes and behaviors of civil servants.

Insufficient degree of disclosure of performance information results to subjects external to the enterprise. At present, the content and scope of performance information disclosed to the public in practice are still inadequate and insufficient. The performance information is mainly transmitted and shared within the enterprise organization, and the relatively high degree of public disclosure is mainly for online evaluation and enterprise satisfaction information, but the degree of disclosure of other information related to people's livelihood is relatively low, as shown in Figure 4.

The selected expert should be an authority in the field of corporate performance evaluation research or an expert scholar with relevant work experience to ensure the accuracy of the evaluation results. In addition, the selected experts should have relevant research or understanding of the new retail innovation ecosystem. The experts should preferably have experience in researching innovation ecosystem projects, and experienced new retail workers should preferably be financial directors or regional directors. The number of experts should be neither too many nor too few. If the number of experts is too large, on the one hand, the cost is relatively high and the feedback period will be extended accordingly; on the other hand, if the opinions of experts are too diverse in the comprehensive processing, not only

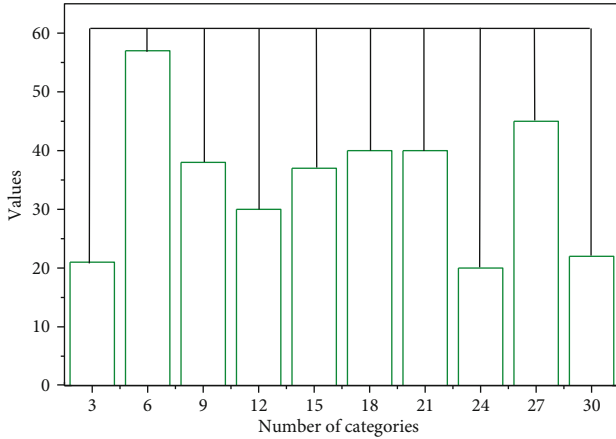


FIGURE 4: Dynamic clustering diagram of factors influencing cross-source multidimensional government data sharing.

is it time-consuming and labor-intensive, but it may not even be possible to reach a unified conclusion. If the number of experts is too small, the analysis results will not only be insufficient for scientific accuracy, but the whole study will be meaningless. Some studies have shown that the number of experts should be best at 7. When the number of experts exceeds 7, the accuracy of the evaluation results does not improve significantly with the number of people, so it is better to choose 7 experts.

## 4. Analysis of Results

**4.1. Results of the Optimized Multidimensional Big Data Matrix Model.** The measurement method of the government recommendation system contains three aspects; one is the comparison of the calculation results of multiple recommendation models. The other is the scoring prediction for result measurement. The recommendation results of a certain enterprise are extracted for manual verification. This study compares the results of enterprise-based recommendation, matter-based recommendation, SVD-based recommendation, and SVD++-based recommendation, with 90% of the original data as training data and 10% of the randomly selected data, as test data, as shown in Figure 5.

By comparing several different algorithms to validate G2B recommendations based on cross-source multidimensional government data, it is appropriate to use the SVD++ algorithm combining enterprise recommendations and matter recommendations as to the core recommendation algorithm. The higher recall rate and accuracy rate in the algorithm comparison indicate that the algorithm is more effective and the lower differential mean evaluation is better. 90% of the original data is used as training data, and 10% of the data is randomly selected. From Figure 5, it can be found that the recall rate and the check accuracy rate are much higher for both the firm-based recommendation and the matter-based recommendation compared to someone, KNN, SVD, and SVD++. Due to the sparsity of the matrix, the enterprise-based recommendation and matter based recommendation have the situation that some enterprises can-

not calculate the results, while SVD++ achieves all enterprise recommendations by dimensionality reduction but its results are less accurate, so these three methods are combined to achieve the complementary and optimized results. Combining enterprise recommendation, matter recommendation, and SVD++ recommendation is a kind of hybrid recommendation combining multiple algorithms, which is to merge and optimize the ranking of multiple calculation results. The algorithm used in this paper is reasonable and effective by comparing the three dimensions of recall rate, check accuracy rate, and differential mean evaluation.

To be able to clearly describe the input-output efficiency of each sector of the enterprise, this study takes each sector of the enterprise as a decision unit in 2020, constructs the CCR model, BCC model, and super-efficiency model, and uses MATLAB software for calculation. By calculating the comprehensive efficiency, pure technical efficiency, scale efficiency, and super efficiency values, the comparative analysis of the input-output efficiency of each enterprise sector in 2020 is carried out, and the calculation results are shown in Figure 6.

The lower the efficiency value, the more serious the waste of resources. In particular, the overall efficiency of the Agriculture Bureau and the Environmental Protection Bureau is lower than 0.7, which is relatively low. The comprehensive efficiency in Figure 6 is calculated by the CCR model, and the results show that the comprehensive efficiency values of Education Bureau, Quality and Technical Supervision Bureau, Public Security Bureau, and Health and Family Planning Bureau are all equal to 1, which means that the DEA of these departments is relatively effective, indicating that these four departments do not have pure technical inefficiency or scale inefficiency, indicating that these departments have achieved effectiveness in terms of technology and scale. This indicates that these sectors are effective in terms of technology and scale and that they can achieve efficient and solid returns to scale with proper inputs in the process of operation. In addition to these four departments, the other three departments: Administration for Industry and Commerce, Bureau of Agriculture, and Environmental Protection Bureau have integrated efficiency values less than 1. They all have different degrees of integrated efficiency inefficiency, indicating that these departments have some degree of resource waste, and some of the input resources cannot be effectively transformed into outputs, and some of the resources invested in the data sharing work do not play the proper value. In particular, the comprehensive efficiency of the Agriculture Bureau and the Environmental Protection Bureau is below 0.7, which is low in relative terms. Low overall efficiency may be caused by managerial errors and unreasonable scale, and technical efficiency can be improved by optimizing resource allocation and improving resource management methods.

According to the calculation of the BCC model, the scale efficiency values of the Bureau of Quality and Technical Supervision, the Bureau of Education, and the Bureau of Health and Family Planning are all lower than reasonable levels, which indicates that there is a waste of resources in the construction process, so it is recommended that these



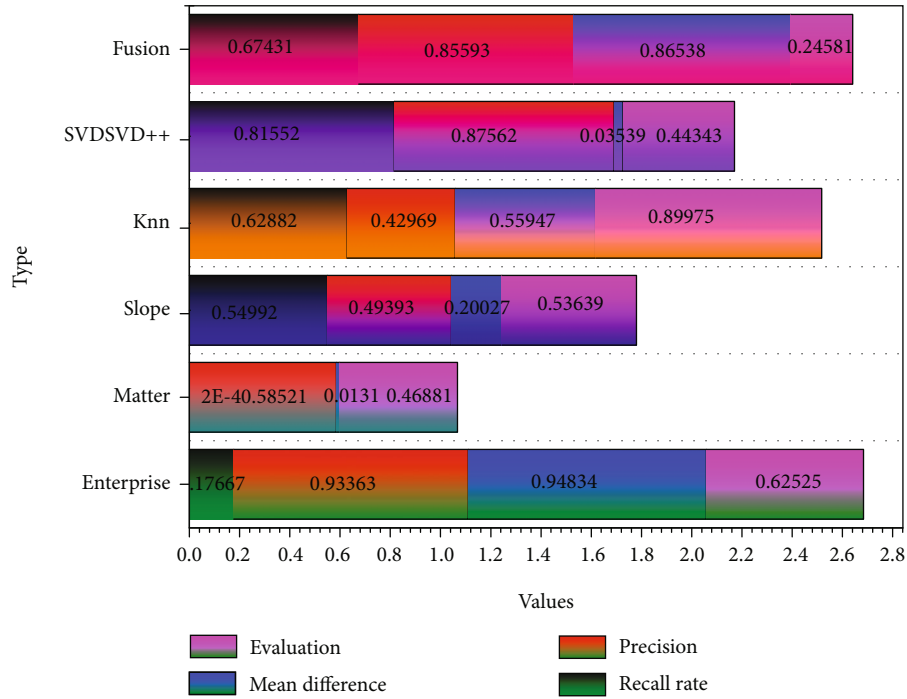


FIGURE 5: Comparison of the efficiency of various algorithms.

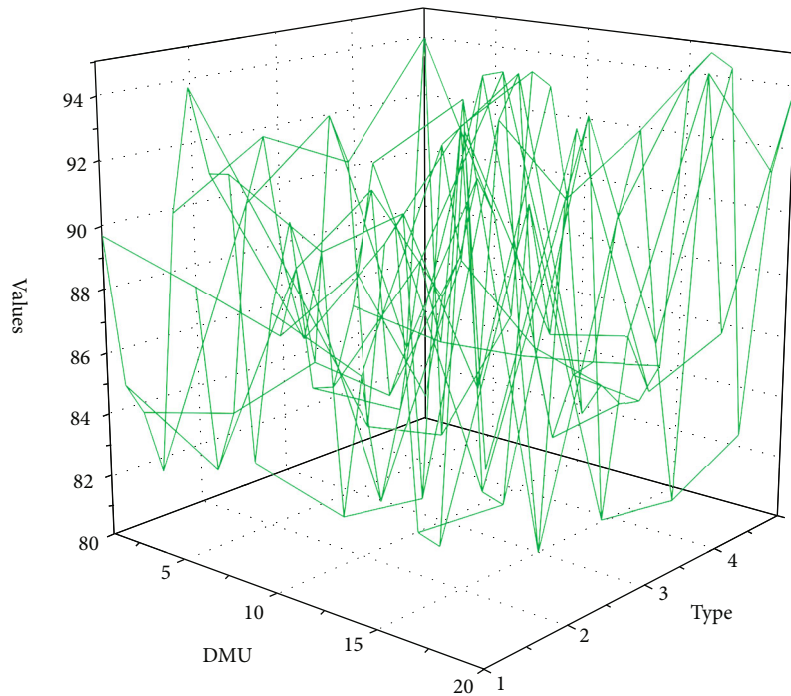


FIGURE 6: Efficiency evaluation value of each department.

units do a good job of managing the whole process in the cross-source multidimensional data sharing work. This includes top-level planning at the early stage of the project, consulting and design at the system implementation stage, and more comprehensive and careful management of the bidding process. We should hire professional consulting

companies, supervisory companies, bidding companies, and software evaluation companies to do a good job and conduct forward-looking research jointly with universities and professional research institutions to promote the sharing of cross-source multidimensional data and actively play the value of data application.

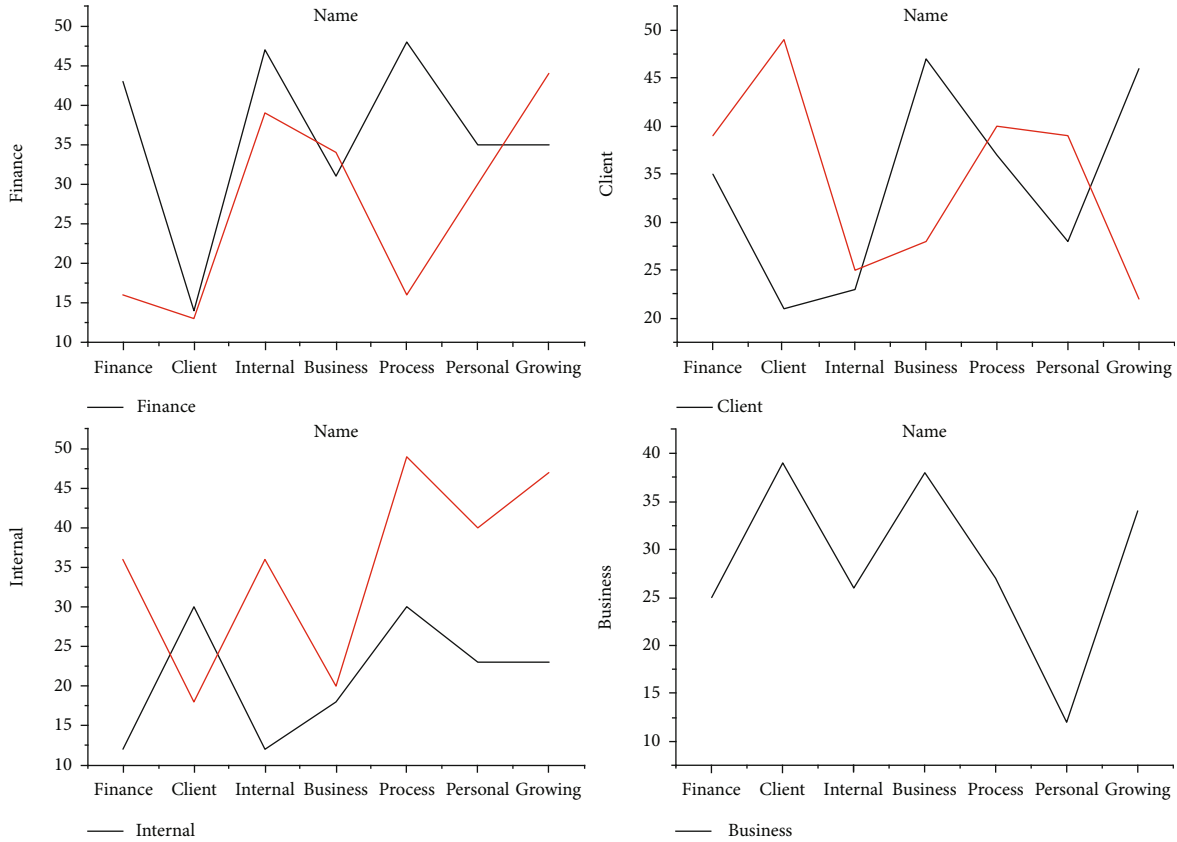


FIGURE 7: Judgment of the weighting factor of the first-level index.

4.2. *Experimental Results.* A two-dimensional table consisting of rows and columns is constructed for the performance indicators of the same level to be evaluated, and all the performance indicators to be evaluated at the same level are listed in the rows and columns, respectively. The following is an example of the construction method of the weight factor judgment table for the first level performance indicators of the data team performance evaluation system. As shown in Figure 7, the first column lists the four levels 1 performance indicators of the data team, which are financial level, customer level, internal business process level, and personal growth level. The first column also lists these four indicators.

Each cell at the intersection of the rows and columns is filled with a score that compares the importance of the row and column indicators to each other. The criteria and range of comparative scores were filled in by an expert panel of 18 experts based on the quantitative scale of importance. For example, if an expert think that the customer-level indicators are less important than the financial-level indicators, he/she fills in the cell where the customer-level and financial-level indicators are intersected with the value 4. When all the tables are filled, the table of the weighting factors of the first-level indicators of an expert is formed. The table of 18 levels 1 index weighting factors is formed. The arithmetic mean of the values in the 18 cells of the table of the weighting factors of the first-level indicators is calculated, and the final table of the weighting factors of the first-level indicators of the data team is formed. It is impor-

tant to note that the 18 experts' scores need to be checked for consistency, and if they pass, the next step can be performed. Similarly, for the four dimensions of the balanced scorecard, all the following second-level indicators are also used in the same way to construct the weighting factor judgment table separately, and the final weighting factor judgment will be formed.

It is also necessary to combine the weights of the indicators of the two tiers, i.e., the indicator weights of the second tier are multiplied with the corresponding weights of the first tier, which is called the total arrangement of weights. The specific calculation method is to multiply the weights of the second tier (C tier) by the respective corresponding weights of the first tier (B tier) to obtain the final team performance index weights. The team performance appraisal results can be analyzed to adjust the pay level of the team. According to the different results of team performance appraisal, it can be divided into the adjustment of overall team salary level and the adjustment of salary level of some positions. The adjustment of the overall salary level of the team is made according to the comprehensive score of the team performance appraisal results. And the adjustment of the salary level of some positions is based on the completion of each performance appraisal index. The principle of salary level adjustment is to first determine the overall performance level of the team according to the total score of the team's performance appraisal results and, then, adjust the salary of some positions according to the completion of each

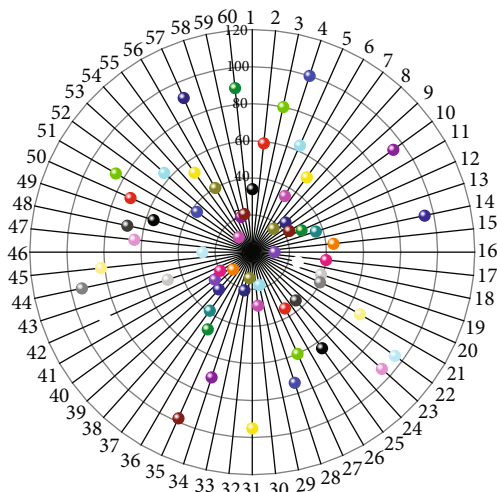


FIGURE 8: Radar chart of performance elements at the same level.

performance appraisal index. The principle of salary level adjustment is to first determine the overall team performance level based on the total score of the team performance appraisal results and, then, to adjust the salary of some positions according to the completion of each performance appraisal index.

The evaluation of R&D personnel should focus on the combination of result-oriented quantitative and qualitative indicators, and the quantitative evaluation should be result-oriented, considering the R&D process, while focusing on the realization of R&D personnel's self-worth, so that the performance of R&D personnel can be truly and reasonably evaluated. For the qualitative part, I can ask experts and relevant personnel to depict a radar chart of performance evaluation elements for R&D personnel's job competency and then conduct a comprehensive analysis by combining big data technology to combine the qualitative part with the quantitative output part, to determine the most reasonable index weights. The radar diagram of R&D personnel performance evaluation is shown in Figure 8.

Risk early warning is a system that monitors changes in the trend of risk factors based on the characteristics of the object of study, collects related information, and measures the strength of deviation from the warning line for each type of risk state and, at the same, time adopts preemptive measures for decision-makers and emits early warning signals. The construction of a risk warning system requires the establishment of an indicator measurement system and in-depth analysis and processing of indicators. The evaluation indicator system in HR performance management is based on performance evaluation indicators and risk indicator evaluation levels, which are defined digitally, and the early warning system indicators are included in the performance indicator design considerations, which is the basis for the establishment of the early warning system.

Second, according to the actual situation of the enterprise, combined with the performance evaluation index, use reasonable early warning data processing method, set early warning processing model, and comprehensive evalua-

tion data index system. The risk model can be set according to the position, work style, and process, generally; there are position risk model, work risk model, process risk model, other models, etc. Finally, according to the degree of completion of performance indicators in the actual work and early warning system indicators to compare, set the early warning interval, when the gap between the actual completion of indicators and early warning indicators exceeds the early warning interval, the early warning system will measure the risk to formulate processing countermeasures and send to the executive and supervisory personnel for the relevant personnel to correct the deviation.

## 5. Conclusion

The positive impact of big data fusion capability on enterprise innovation performance has not been verified. To explore the reason, we can return to the definition of "big data convergence capability", which is the orderly integration of specific business and big data systems, which will break the traditional management style, redeploy and optimize enterprise resources, enhance business process sharing and interactivity, and thus improve the efficiency of enterprise operation. Realizing the orderly combination of business and big data systems, as well as the redeployment and optimization of enterprise resources is a gradual process, which cannot be completed in a short time. At present, China's big data industry is in the primary development stage, especially the relevant technology is in the stage of technical barriers to overcome, and the external conditions for cultivating the enterprise's big data integration capability have not reached a mature time. Therefore, the positive impact of big data integration capability on the innovation performance of enterprises has not been revealed yet. Big data strategic collaboration capability is a positive adjustment of enterprises' strategies in the new era of change. In the fierce market competition, big data strategic collaboration capability enhances the adaptability of enterprises to the market environment, and the actual needs of customers support the concrete implementation of enterprise strategies based on the guidance of learning mechanisms and promote the management innovation activities of enterprise management members from the perspective of internal management and operation. Big data strategic collaboration capability is an important path for the transformation of enterprise innovation performance, providing support for the future development strategy of the enterprise and realizing the improvement of the overall competitiveness of the enterprise.

## Data Availability

The data used to support the findings of this study are included within the article.

## Conflicts of Interest

The author does not have any possible conflicts of interest.

## References

- [1] Z. H. Munim, M. Dushenko, V. J. Jimenez, M. H. Shakil, and M. Imset, "Big data and artificial intelligence in the maritime industry: a bibliometric review and future research directions," *Maritime Policy & Management*, vol. 47, no. 5, pp. 577–597, 2020.
- [2] J. Wen, J. Yang, B. Jiang, H. Song, and H. Wang, "Big data driven marine environment information forecasting: a time series prediction network," *IEEE Transactions on Fuzzy Systems*, vol. 29, no. 1, pp. 4–18, 2020.
- [3] H. Patel, D. Prajapati, D. Mahida, and M. Shah, "Transforming petroleum downstream sector through big data: a holistic review," *Journal of Petroleum Exploration and Production Technology*, vol. 10, no. 6, pp. 2601–2611, 2020.
- [4] X. Zhu and Y. Yang, "Big data analytics for improving financial performance and sustainability," *Journal of Systems Science and Information*, vol. 9, no. 2, pp. 175–191, 2021.
- [5] V. Kumar and D. Ramachandran, "Developing firms' growth approaches as a multidimensional decision to enhance key stakeholders' wellbeing," *International Journal of Research in Marketing*, vol. 38, no. 2, pp. 402–424, 2021.
- [6] S. A. Gawankar, A. Gunasekaran, and S. Kamble, "A study on investments in the big data-driven supply chain, performance measures and organisational performance in Indian retail 4.0 context," *International Journal of Production Research*, vol. 58, no. 5, pp. 1574–1593, 2020.
- [7] S. Thudumu, P. Branch, J. Jin, and J. Singh, "A comprehensive survey of anomaly detection techniques for high dimensional big data," *Journal of Big Data*, vol. 7, no. 1, 2020.
- [8] S. Chehbi-Gamoura, R. Derrouiche, D. Damand, and M. Barth, "Insights from big Data Analytics in supply chain management: an all-inclusive literature review using the SCOR model," *Production Planning & Control*, vol. 31, no. 5, pp. 355–382, 2020.
- [9] F. Mohammadi and M. Sahraei-Ardakani, "Multidimensional scenario selection for power systems with stochastic failures," *IEEE Transactions on Power Systems*, vol. 35, no. 6, pp. 4528–4538, 2020.
- [10] L. Ismail and H. Materwala, "Computing server power modeling in a data center," *ACM Computing Surveys (CSUR)*, vol. 53, no. 3, pp. 1–34, 2020.
- [11] Z. Liu, N. Wu, Y. Qiao, and Z. Li, "Performance evaluation of public bus transportation by using DEA models and Shannon's entropy: an example from a company in a large city of China," *IEEE/CAA Journal of Automatica Sinica*, vol. 8, no. 4, pp. 779–795, 2020.
- [12] X. Zhou, W. Liang, K. I.-K. Wang, and L. T. Yang, "Deep correlation mining based on hierarchical hybrid networks for heterogeneous big data recommendations," *IEEE Transactions on Computational Social Systems*, vol. 8, no. 1, pp. 171–178, 2020.
- [13] Q. Xiao, M. Shan, M. Gao, X. Xiao, and H. Guo, "Evaluation of the coordination between China's technology and economy using a grey multivariate coupling model," *Technological and Economic Development of Economy*, vol. 27, no. 1, pp. 24–44, 2021.
- [14] W. Höpken, T. Eberle, M. Fuchs, and M. Lexhagen, "Improving tourist arrival prediction: a big data and artificial neural network approach," *Journal of Travel Research*, vol. 60, no. 5, pp. 998–1017, 2021.
- [15] J. Zhang, Y. Liu, H. Liu, and J. Wang, "Learning local–global multiple correlation filters for robust visual tracking with Kalman filter redetection," *Sensors*, vol. 21, no. 4, p. 1129, 2021.
- [16] Y. Jiang, X. Gu, D. Wu, W. Hang, J. Xue, and S. Qiu, "A novel negative-transfer-resistant fuzzy clustering model with a shared cross-domain transfer latent space and its application to brain CT image segmentation," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 18, no. 1, pp. 40–52, 2021.
- [17] M. Gao, W. Cai, and R. Liu, "AGTH-net: attention-based graph convolution-guided third-order hourglass network for sports video classification," *Journal of Healthcare Engineering*, vol. 2021, 2021.
- [18] L. Perkhofer, W. Conny, and P. Hofer, "Does design matter when visualizing Big Data? An empirical study to investigate the effect of visualization type and interaction use," *Journal of Management Control*, vol. 31, no. 1-2, pp. 55–95, 2020.
- [19] Z. Wang, P. Zhang, W. Sun, and D. Li, "Application of data dimension reduction method in high-dimensional data based on single-cell 3D genomic contact data," *ASP Transactions on Computers*, vol. 1, no. 2, 2021.
- [20] C. Zhao, L. Ren, Z. Zhang, and Z. Meng, "Master data management for manufacturing big data: a method of evaluation for data network," *World Wide Web*, vol. 23, no. 2, pp. 1407–1421, 2020.
- [21] W. Cai, Y. Song, and Z. Wei, "Multimodal data guided spatial feature fusion and grouping strategy for E-commerce commodity demand forecasting," *Mobile Information Systems*, vol. 2021, 2021.
- [22] H. Li, X. Bu, X. Liu et al., "Evaluation and prediction of blast furnace status based on big data platform of ironmaking and data mining," *ISIJ International*, vol. 61, no. 1, pp. 108–118, 2021.
- [23] A. Li and R. C. Bergan, "Clinical trial design: past, present, and future in the context of big data and precision medicine," *Cancer*, vol. 126, no. 22, pp. 4838–4846, 2020.
- [24] C. Yan, G. Pang, X. Bai et al., "Beyond triplet loss: person re-identification with fine-grained difference-aware pairwise loss," *IEEE Transactions on Multimedia*, p. 1, 2021.
- [25] W. Sun, P. Zhang, Z. Wang, and D. Li, "Prediction of cardiovascular diseases based on machine learning," *ASP Transactions on Internet of Things*, vol. 1, no. 1, pp. 30–35, 2021.
- [26] Q. G. Shao, J. J. H. Liou, S. S. Weng, and P. Su, "Constructing an entrepreneurship project evaluation system using a hybrid model," *Journal of Business Economics and Management*, vol. 21, no. 5, pp. 1329–1349, 2020.

## Research Article

# Application of Flipped Classroom in the Era of Big Data: What Factors Influence the Effect of Teacher-Student Interaction in Oral English Teaching

Yan Liu <sup>1</sup> and Wenjin Qi<sup>2</sup>

<sup>1</sup>Department of Foreign Languages, Yuncheng University, Yuncheng Shanxi 044000, China

<sup>2</sup>Yuncheng University, Yuncheng Shanxi 044000, China

Correspondence should be addressed to Yan Liu; [liuyan@ycu.edu.cn](mailto:liuyan@ycu.edu.cn)

Received 6 July 2021; Revised 17 July 2021; Accepted 26 July 2021; Published 26 August 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Yan Liu and Wenjin Qi. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Teachers now have a new teaching model to refer to with the emergence of the flipped classroom. The teacher-student interaction is thought to be influenced by teaching mode as a key indicator of improving teaching efficiency. We develop a research model based on a literature review to explain the drawbacks of teacher-student interaction in flipped classrooms with oral English instruction. SmartPLS software was used to analyze the survey data of 84 students from three classes. Emotional resistance has a significant negative impact on the effect of teacher-student interaction in the flipped classroom, whereas learning autonomy and video content quality have a significant positive impact. Our research adds to the body of knowledge in the field of flipped classrooms and provides theoretical support for teachers in colleges and universities who want to use this teaching method.

## 1. Introduction

As an important part of teaching for college teachers, teacher-student interaction has a significant impact on the teaching quality of teachers and the leaning efficiency of students [1]. Appropriate and high-level teacher-student interaction can mobilize students' enthusiasm in class and guide students to think and express their views better. On the contrary, it may make the class boring and then affect the learning efficiency of students [2].

Oral English teaching, as an applied and practical course, requires relatively high interaction between teachers and students [3]. As an oral English teacher, it is not only necessary to get students to speak English and interact with classmates and teachers but more importantly, to get them actively involved in the classroom teaching and discussion. In other words, we need to get students to take the initiative, actively express their ideas, and participate in the class.

The arrival of the era of big data provides a solution to the issue of teacher-student interaction in oral English teaching [4]. "Flipped classroom" is a typical solution of teacher-student

interaction. Flipped classrooms typically rearrange the time in and out of the classroom, transferring learning decisions from teachers to students [5]. Under this teaching mode, students can arrange their study time better. For example, students use extracurricular time to prepare questions for class in advance, so that they can actively and effectively interact with teachers and other students during class. Moreover, in the Internet era, students can learn a variety of online courses through the Internet without having to go to school to be taught by teachers, while the advent of the era of big data provides students with a wealth of online materials. Students can obtain rich learning materials independently through the Internet under the mode of "flipped classroom."

Recently, artificial intelligence as a hot research field has attracted the attention and application of many scholars in many fields [6–8]. As a product of the era of artificial intelligence [9–12], big data [13], and machine learning [14], flipped classroom has attracted the attention of academia and higher education circles as soon as its concept and model were launched. For example, Herreid and Schiller [15] have discussed the positive and negative aspects of flipped classrooms.

Herreid et al. [16] believed that the teaching mode of flipped classroom provides a very good choice for teachers who adopt the case study method in teaching. However, Abdullah et al. [17] proposed that the application of flipped classroom mode in oral English teaching may cause the anxiety of learners. This point of view provides new inspiration for our study. Instead, Abdullah et al. [17] were thinking about the drawbacks of flipped classroom, a widely accepted teaching method, that might hinder students' learning.

The viewpoints of the above scholars have caught our attention. We cannot help thinking, what are the potential disadvantages factors of flipped classroom in oral English teaching that will affect the interaction between teachers and students? Therefore, following this idea, we mainly discuss the possible drawbacks of flipped classroom in oral English teaching and puts forward feasible solutions. In this study, we constructed a model of the factors influencing the disadvantages of flipped classroom in oral English teaching. Our study contribution is to enrich the research literature of flipped classroom and expand the teaching practice of oral English. Meanwhile, the exploration of adverse factors is conducive to providing practical implications for teachers of oral English teaching and preventing students from finding it difficult to adapt to flipped classroom.

The rest of the paper is arranged as follows: the second section presents the literature review; the third section proposes the research model and hypothesis; the fourth section provides the method and data results; and finally, the fifth section proposes the conclusions, limitations, and the future work.

## 2. Literature Review

**2.1. Definition of Flipped Classroom.** It is generally believed in the academia that Lage et al. [18] were the first researchers to put forward the idea of "flipped classroom," but they did not clearly define this concept [19]. Until 2007, Bergmann and Sams, chemistry teachers at Wood Land Park High School in Colorado, USA, began to use video software to record PPT and attach the voice of explanation [17]. The videos they record are uploaded to the Internet to make up for absent students. Soon, they were experimenting with something more groundbreaking, gradually based on students watching videos and listening to lectures at home; in class, teachers mainly help students with questions or provide help to students who have difficulties in doing experiments. The mode has since become popular in the United States.

Therefore, in a sense, flipped classroom is different from the traditional teaching method. In this mode, teachers give the teaching content in the classroom to students to learn independently after class; during class time, teachers and students interact with each other to discuss questions that students do not understand or to put the original homework in class for discussion [5].

**2.2. Potential Disadvantages of Flipped Classroom in Oral English Teaching.** In oral English teaching, flipped classroom has many advantages. For example, it provides a space for students to study independently, which is beneficial for students to make full preparation for the learning content after

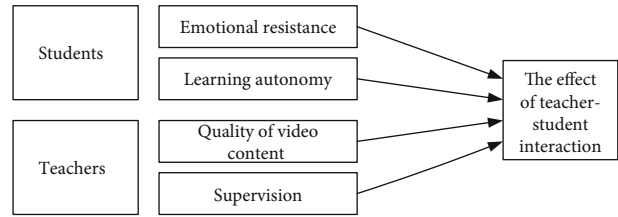


FIGURE 1: Research model.

class, and the interaction between students and teachers in class is more efficient [20]. The existence of these advantages attracts many teachers and students to try to learn through this teaching method.

However, we believe that the potential drawbacks of flipped classroom in oral English teaching should be paid more attention to by teachers and students. Bishop and Verleger [21] found that students preferred interactive classes to video lectures. Herreid et al. [5] found in their research that flipped classroom has potential traps. For example, students may be resistant at the beginning of learning; the videos prepared by teachers for students are all unified, so it is difficult to make differentiated customization; and the poor quality of instructional videos may also be a potential disadvantage. In addition, Critz and Knight [22] also raised concerns about this teaching method. One of the main concerns is that it is difficult to monitor and control the learning process and learning efficiency of students.

Existing literature provides a rich foundation for us to study flipped classroom. In view of the disadvantages of flipped classroom in oral English teaching, this paper attempts to investigate and verify these potential factors. The aims are to clearly explain these potential disadvantages and propose the workable solutions.

## 3. Theoretical Model Construction and Hypothesis

**3.1. Theoretical Model.** Based on literature analysis and from the perspective of "teacher-student interaction," we constructed a model of the factors influencing the disadvantages of flipped classroom in oral English teaching. As shown in Figure 1, the model identifies four factors that may affect "teacher-student interaction," including students' emotional resistance, learning autonomy, quality of video content, and teacher supervision. Among them, the first two factors mainly come from the students; the last two factors come from the teachers.

### 3.2. Hypothesis

**3.2.1. Factors from Students.** First is emotional resistance. In oral English teaching, compared with the traditional teaching method, flipped classroom may cause students' emotional resistance due to the difficulty of data query, long time to solve problems, and difficulty to adapt to the new teaching mode [23]. For example, when students are learning by video, they cannot interact with the teacher directly if they do not understand the issue. Instead, they need to search the information

TABLE 1: Measurement scale.

Latent variables	Observed variables	References
Emotional resistance (ER)	1a: I am resistant to flipped classrooms.	Hew and Lo [23]; Huseyin et al. [24]
	1b: This kind of emotional resistance will slow down my learning efficiency.	
	1c: This emotional resistance made me perform poorly in the teacher-student interaction.	
Learning autonomy (LA)	2a: In my opinion, flipped classroom requires high learning autonomy of students.	Sergis et al. [25]; Zheng et al. [26]
	2b: In the process of video learning in flipped classroom, I need to maintain active learning autonomy all the time.	
	2c: If I have less autonomy in learning, it will not be good for me to interact with the teacher in class.	
Quality of video content (QVC)	3a: I think the quality of video content seriously affects my learning efficiency.	Bishop and Verleger [21]; Pickering and Roberts [27]
	3b: I hope teachers pay attention to improving the quality of teaching videos.	
	3c: I think the high quality of teaching videos is conducive to my interaction with teachers.	
Supervision (S)	4a: In my opinion, teachers cannot effectively supervise us under the flipped classroom.	Chen et al. [28]
	4b: I think the lack of supervision of learning is not conducive to the improvement of learning efficiency.	
	4c: In my opinion, unsupervised learning is not conducive to teacher-student interaction in the classroom.	
The effect of teacher-student interaction (ETSI)	a: I think the flipped classroom provides a good platform for teacher-student interaction.	Bunning et al. [1]
	b: I attach great importance to the teacher-student interaction in the flipped classroom.	
	c: I think teacher-student interaction is a good way to improve learning efficiency.	

TABLE 2: Descriptive statistical results of the investigated students.

Control variables	Items	Rate	Control variables	Items	Rate
Age	18-20	5%	Attitude towards flipped classroom	Support	47%
	20-25	92%		Neither for nor against	7%
	>25	3%		Oppose	46%
Gender	Male	46%	Duration in flipped classroom	≤1 year	0%
	Female	54%		>1 year	100%

by themselves to solve the questions in their minds. However, students may not know where to look for relevant information. In addition, when students study video courseware by themselves, they can only think by themselves when they encounter questions, so it is difficult for them to get help from teachers and classmates in a short time. Moreover, for a long time, the traditional teaching method has made students adapt well. When the new flipped classroom is implemented, students may have the problem of not adapting to the transformation of learning mode [24]. Students' emotional resistance due to the above reasons will not be conducive to the effect of teacher-student interaction in flipped classroom. Based on the above analysis, we propose the following hypotheses:

TABLE 3: Results of reliability and validity tests.

Construct	CA	CR	AVE
ER	0.935	0.956	0.879
LA	0.713	0.728	0.643
QVC	0.700	0.732	0.655
S	0.716	0.752	0.703
ETSI	0.816	0.891	0.731

ER: emotional resistance; LA: learning autonomy; QVC: quality of video content; S: supervision; ETSI: the effect of teacher-student interaction.

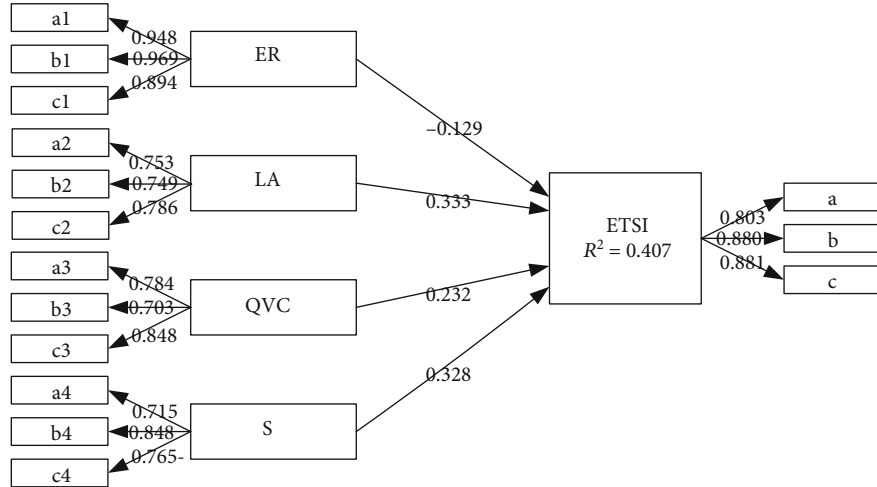


FIGURE 2: Model calculation results.

H1. The greater the emotional resistance of students, the worse the effect of teacher-student interaction in the flipped classroom.

Second is learning autonomy. Learning autonomy refers to the degree to which a student actively learns based on a video course outside of the classroom [25]. Compared with the traditional in-class teaching mode, the knowledge learning of flipped classroom largely depends on students' learning of videos after class. They need to be self-controlled to complete the learning of the video content without the supervision of the teacher. In addition, if students encounter some learning questions, they also need to independently inquire relevant materials and conduct independent understanding [26]. Therefore, the higher the learning autonomy of students is, the more beneficial the activity effect of teachers and students in the flipped classroom will be. Therefore, we propose the following hypothesis:

H2. The stronger the students' learning autonomy, the better the effect of teacher-student interaction in the flipped classroom.

**3.2.2. Factors from Teachers.** First is quality of video content. As stated by Bishop and Verleger [21], students sometimes prefer interactive classes to video lectures. One of the important reasons for this is that the quality of video content may affect students' learning enthusiasm [27]. If it is a high quality video learning course, it may attract students to learn actively. However, after the flipped classroom mode is implemented, teachers need to do more lesson preparation, one of which is recording video lectures. This poses some challenges for teachers who have been working in the classroom for a long time. Once the quality of the video course cannot meet the requirements of students, students' learning enthusiasm will decline and even affect the learning effect of students. The resulting consequences will further affect the teacher-student interaction in flipped classroom. Therefore, we propose the following hypothesis:

TABLE 4: Average variance extracted (AVE) square root and factor correlation coefficient.

	<i>H</i>	H1	H2	H3	H4
ETSI	0.855				
ER	0.082	0.938			
LA	0.512	0.119	0.687		
QVC	0.407	0.296	0.361	0.696	
<i>S</i>	0.465	0.315	0.336	0.284	0.674

*Italic elements are the square root of average variance extracted (AVE).*

H3. The higher the quality of the video content provided by the teacher, the better the effect of the teacher-student interaction in the flipped classroom.

Second is supervision. In flipped classroom, it is difficult for teachers to supervise students' learning progress and learning effect [28]. This is because the content of the course needs to be done independently by students outside of the classroom. What the teacher can do is to assist the students in the class to solve their questions. In fact, in the traditional classroom teaching mode, teachers can evaluate the learning effect of students through their listening state and the number of opinions expressed [29]. However, in the flipped classroom, if the teacher's supervision is lacking, students may not be able to achieve the expected video learning effect. This will affect the interaction between teachers and students later. Therefore, we propose the following hypothesis:

H4. The greater the difficulty of teacher supervision, the worse the effect of teacher-student interaction in the flipped classroom.

## 4. Methodology and Empirical Results

**4.1. Questionnaire Design.** In order to verify the above research hypothesis, we selected three classes of oral English teaching, a total of 84 students, to conduct a questionnaire



TABLE 5: Results of hypothesis testing.

Hypothesis	Path	Path coefficient	<i>t</i> -value	<i>P</i> value	Hypothesis supported?
H1	ER → ETSI	-0.129	3.080	0.002	Y
H2	LA → ETSI	0.333	1.987	0.046	Y
H3	QVC → ETSI	0.232	2.786	0.001	Y
H4	S → ETSI	0.328	1.392	0.165	N

survey. First of all, we designed the questionnaire. The questionnaire information consists of two parts. The first part is the basic information of the investigated students, including their age, gender, attitude towards flipped classroom, and the duration in flipped classroom. The second part is four latent variables. We used 5-point Likert scale to quantify the observed variables of latent variables [30]. All observed variables were designed according to the views of the existing literature, as shown in Table 1.

In order to prevent the influence of potential factors on the results, we took the basic information of the investigated students, namely, age, gender, attitude towards flipped classroom, and the duration in flipped classroom, as the control variables.

**4.2. Data Collection.** After designing the questionnaire, we gave it to four students in the oral English course for evaluation and adjusted it to some extent according to their feedback. After that, we distributed questionnaires in three classes of oral English. All questionnaires were filled out anonymously by the students. Questionnaires were distributed online and offline. Our survey began in October 2020 and ended in November of the same year. All the students participated voluntarily and were not paid. Fortunately, all 84 students actively completed the questionnaire, and there was no invalid questionnaire. The descriptive statistical results of this questionnaire survey are shown in Table 2.

**4.3. Data Analysis.** In this paper, SmartPLS software [31] is used to analyze the collected data. SmartPLS software has very good adaptability in building structural equation model, testing the reliability and validity of the scale, and testing the research hypothesis [32]. In addition, the software can be used for multifactor causality studies [33].

**4.3.1. Reliability and Validity Tests.** As shown in Table 3, Cronbach's alpha (CA), combined reliability (CR), and average extraction variance (AVE) values of the five latent variables all reached the relevant threshold [34, 35]. Meanwhile, in Figure 2, the external load values for each of the observed variables are greater than 0.7. This indicates that the scale designed and the model constructed by us have very good reliability and validity [34]. In addition, as shown in Table 4, the square root of AVE is greater than the correlation coefficient between it and other potential variables, indicating that there is no multicollinearity among these 5 latent variables [36].

**4.3.2. Hypothesis Testing.** SmartPLS software was used to verify the proposed research hypothesis, and the results are shown in Table 5. The final results showed that hypotheses

H1, H2, and H3 passed the hypothesis test ( $P \leq 0.05$ ), while H4 failed the hypothesis test ( $P \text{ value} > 0.05$ ) [37]. The  $R^2$  of the constructed model is 0.407, which indicates that the constructed model has good explanatory ability [37]. In addition, the SRMR value of the model is 0.043, less than 0.06. Therefore, our PLS-SEM has good model fit.

As can be seen from the results of hypothesis testing, hypotheses H1, H2, and H3 are consistent with existing literature and viewpoints, wherein H1 is consistent with the research conclusions of Hew and Lo [23] and Huseyin et al. [24]; that is, in flipped classroom, the emotional resistance of students is inversely proportional to the effect of teacher-student interaction. H2 has the same opinion as Sergis et al. [25] and Zheng et al. [26]; that is, in flipped classroom, students' learning autonomy is in direct proportion to the effect of teacher-student interaction. H3 is consistent with the research results of Bishop and Verleger [21] and Pickering and Roberts [27]; that is, the higher the quality of video content provided by teachers, the better the teacher-student interaction in flipped classroom. However, hypothesis H4 has not been tested. This indicates that the research results do not support the viewpoints of Chen et al. [28].

## 5. Conclusions

The creation of a flipped classroom offers teachers a new and worthy-of-reference method of teaching. Many university teachers are therefore trying to use this mode of teaching every day. Teacher-student interaction is considered to be an important indicator of improving teaching efficiency. This paper examines the potential disadvantages of flipped classroom mode, based on the situation with oral English instruction. We use a literature review to build a research model to see if these flaws are real and have an impact on teacher-student interaction. Emotional resistance, learning autonomy, video content quality, and supervision are four unfavorable factors that affect teacher-student interaction. Our findings show that emotional resistance has a significant negative impact on the effect of teacher-student interaction in the flipped classroom, whereas learning autonomy and video content quality have a significant positive impact, according to a questionnaire survey and empirical analysis. Furthermore, supervision has no bearing on the situation.

Our research adds to the body of knowledge in the field of flipped classrooms and provides theoretical support for teachers in colleges and universities who want to implement this teaching method. Meanwhile, we have created a model of the factors that influence the flipped classroom's disadvantages in oral English instruction. This model can be used as a theoretical framework for discussing other influencing

factors in this research area. It is worth noting that, while the flipped classroom mode has many benefits, it is important to focus on overcoming students' emotional resistance. For example, before implementing the flipped classroom mode, teachers can provide students with the necessary training and a question-and-answer channel for after-class learning. In addition, students' autonomous learning abilities should be cultivated, and the quality of video courses should be improved as much as possible. It is important to note that spoken English is a subject with a wide range of applications. The key to improving students' oral skills is for them to study and practice after class. Students' independent learning is expected to be high in a flipped classroom. This concept is in line with improving students' ability to participate in active learning. Meanwhile, as video content quality improves, students will be more likely to actively participate in flipped classroom learning. Teachers can thus combine fun with the applicability of teaching content when recording videos.

This paper has a few drawbacks. For starters, our sample size is small, which may have an impact on the universality of our findings. As a result, in the future, the sample size must be increased and the relevant research findings must be improved. Second, as information and communication technology advances, so does the specific application of flipped classroom, and some potential disadvantages may not be fully considered in this paper. For example, the speed of online transmission over the network may have an impact. Finally, we discuss the potential drawbacks of flipped classrooms from the viewpoints of teachers and students. External perspectives, such as college and even the living environment, are also worthwhile to investigate. Finally, because oral English is a subject with broad application, it should be investigated further in the flipped classroom teaching mode to see if it has a different effect and applicability than other courses. Furthermore, the flipped classroom mode has numerous advantages, with the most appealing factors for students deserving of further investigation in the future.

### Data Availability

The data used to support the findings of this study are included within the article.

### Conflicts of Interest

The authors declare that they have no conflicts of interest.

### Acknowledgments

The study was supported by Shanxi Scholarship Council of China (Grant No. 2016-107).

### References

- [1] K. Bunning, B. Heath, and A. Minnion, "Interaction between teachers and students with intellectual disability during computer-based activities: the role of human mediation," *Technology and Disability*, vol. 22, no. 1-2, pp. 61–71, 2010.
- [2] N. R. Santoso, M. Nombrado, M. T. De Guzman, S. D. Yumul, and R. M. Mariano, "Teachers' professional identity construction on Facebook using the teacher-student interaction perspective," *Jurnal Studi Komunikasi (Indonesian Journal of Communications Studies)*, vol. 5, no. 1, p. 1, 2021.
- [3] G. Minghe and W. Yuan, *Affective Factors in Oral English Teaching and Learning*, Higher Education of Social Science, 2013.
- [4] L. Liang, Q. Yin, and C. Shi, "Exploring proper names online and its application in English teaching in university," *ASP Transactions on Computers*, vol. 1, no. 1, pp. 24–29, 2021.
- [5] C. F. Herreid, N. A. Schiller, K. F. Herreid, and C. B. Wright, "Case study: a chat with the survey monkey: case studies and the flipped classroom," *Journal of College Science Teaching*, vol. 44, no. 1, pp. 75–80, 2014.
- [6] M. Gao, W. Cai, and R. Liu, "AGTH-Net: attention-based graph convolution-guided third-order hourglass network for sports video classification," *Journal of Healthcare Engineering*, vol. 2021, 10 pages, 2021.
- [7] Y. Jiang, X. Gu, D. Wu, W. Hang, J. Xue, and S. Qiu, "A novel negative-transfer-resistant fuzzy clustering model with a shared cross-domain transfer latent space and its application to brain CT image segmentation," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 18, no. 1, pp. 40–52, 2021.
- [8] S. Qi, X. Ning, G. Yang et al., "Review of multi-view 3D object recognition methods based on deep learning," *Displays*, vol. 69, p. 102053, 2021.
- [9] Y. Tong, L. Yu, S. Li, J. Liu, H. Qin, and W. Li, "Polynomial fitting algorithm based on neural network," *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 1, pp. 32–39, 2021.
- [10] X. Ning, Y. Wang, W. Tian, L. Liu, and W. Cai, "A biomimetic covering learning method based on principle of homology continuity," *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 1, pp. 9–16, 2021.
- [11] W. Cai, Y. Song, and Z. Wei, "Multimodal data guided spatial feature fusion and grouping strategy for E-commerce commodity demand forecasting," *Mobile Information Systems*, vol. 2021, 14 pages, 2021.
- [12] Y. Zhang, Y. Jiang, L. Qi, M. Z. A. Bhuiyan, and P. Qian, "Epilepsy diagnosis using multi-view & multi-medoid entropy-based clustering with privacy protection," *ACM Transactions on Internet Technology*, vol. 21, no. 2, pp. 1–20, 2021.
- [13] Z. Wang, P. Zhang, W. Sun, and D. Li, "Application of data dimension reduction method in high-dimensional data based on single-cell 3D genomic contact data," *ASP Transactions on Computers*, vol. 1, no. 2, pp. 1–6, 2021.
- [14] W. Chu, "Studies on the effects of wiring density on chip package interaction and design optimization with machine learning," Academic Dissertation, 2021.
- [15] C. F. Herreid and N. A. Schiller, "Case study: case studies and the flipped classroom," *Journal of college science teaching*, vol. 42, no. 5, pp. 62–67, 2013.
- [16] C. F. Herreid and N. A. Schiller, "Case studies and the flipped classroom," *Journal of College Science Teaching*, vol. 42, no. 5, pp. 62–66, 2013.
- [17] M. Y. Abdullah, S. Hussin, and K. Ismail, "Does flipped classroom model affect EFL learners' anxiety in English speaking performance?," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 16, no. 1, pp. 94–107, 2021.

- [18] M. J. Lage, G. J. Platt, and M. Treglia, "Inverting the classroom: a gateway to creating an inclusive learning environment," *The Journal of Economic Education*, vol. 31, no. 1, pp. 30–43, 2000.
- [19] J. Bergmann and A. Sams, *Flip Your Classroom: Reach Every Student in Every Class Every Day* June 2021, <http://www.ascd.org/Publications/Books/Overview/Flip-Your-Classroom.aspx>.
- [20] L. Khofifah, N. Supriadi, and M. Syazali, "Model flipped classroom dan discovery learning terhadap kemampuan pemahaman konsep dan pemecahan masalah matematis," *PRISMA*, vol. 10, no. 1, p. 17, 2021.
- [21] J. L. Bishop and M. A. Verleger, "The flipped classroom: a survey of the research," in *2013 ASEE Annual Conference & Exposition*, Atlanta, 2013.
- [22] C. M. Critz and D. Knight, "Using the flipped classroom in graduate nursing education," *Nurse Educator*, vol. 38, no. 5, pp. 210–213, 2013.
- [23] K. Hew and C. Lo, "Flipped classroom improves student learning in health professions education: a meta-analysis," *BMC Medical Education*, vol. 18, no. 1, p. 38, 2018.
- [24] Z. H. Huseyin, K. Sezer, and O. Fezile, "Effects of the gamification supported flipped classroom model on the attitudes and opinions regarding game-coding education," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 13, no. 1, p. 109, 2018.
- [25] S. Sergis, D. G. Sampson, and L. Pelliccione, "Investigating the impact of flipped classroom on students' learning experiences: a self-determination theory approach," *Computers in Human Behavior*, vol. 78, pp. 368–378, 2018.
- [26] M. Zheng, C. C. Chu, Y. J. Wu, and W. Gou, "The mapping of on-line learning to flipped classroom: small private online course," *Sustainability*, vol. 10, no. 3, p. 748, 2018.
- [27] J. D. Pickering and D. J. Roberts, "Flipped classroom or an active lecture?," *Clinical Anatomy*, vol. 31, no. 1, pp. 118–121, 2018.
- [28] K. S. Chen, L. Monrouxe, Y. H. Lu et al., "Academic outcomes of flipped classroom learning: a meta-analysis," *Medical Education*, vol. 52, no. 9, pp. 910–924, 2018.
- [29] A. Pardo, D. Gašević, J. Jovanovic, S. Dawson, and N. Mirriahi, "Exploring student interactions with preparation activities in a flipped classroom experience," *IEEE Transactions on Learning Technologies*, vol. 12, no. 3, pp. 333–346, 2018.
- [30] R. Chomeya, "Quality of psychology test between Likert scale 5 and 6 points," *Journal of Social Sciences*, vol. 6, no. 3, pp. 399–403, 2010.
- [31] G. S. Hubona, "Structural equation modeling (SEM) using SmartPLS software: analyzing path models using partial least squares (PLS) based SEM," in *Proceedings of the 15th Americas Conference on Information Systems, AMCIS 2009*, San Francisco, California, USA, 2009.
- [32] T. Ramayah, J. Cheah, F. Chuah, H. Ting, and M. A. Memon, *Partial Least Squares Structural Equation Modeling (PLS-SEM) Using SmartPLS 3.0: An Updated and Practical Guide to Statistical Analysis*, 2016.
- [33] A. Kamis, R. A. Saibon, F. A. Yunus, M. B. Rahim, L. M. Herrera, and P. L. Montenegro, "The SmartPLS analyzes approach in validity and reliability of graduate marketability instrument," *Social Psychology of Education*, vol. 57, no. 8, pp. 987–1001, 2020.
- [34] G. Prayitno, M. I. Ashari, and W. I. Rukmi, "Structural equation model with partial least square (SEM-PLS) of place dependence with land used change," *Journal of International Studies*, vol. 14, no. 1, 2021.
- [35] H. Bin, F. Zhao, G. Xie et al., "Crowd-sourcing a way to sustainable urban logistics: what factors influence enterprises' willingness to implement crowd logistics?," *IEEE Access*, vol. 8, pp. 225064–225075, 2020.
- [36] M. H. Graham, "Confronting multicollinearity in ecological multiple regression," *Ecology*, vol. 84, no. 11, pp. 2809–2815, 2003.
- [37] R. Wetzels and E. J. Wagenmakers, "A default Bayesian hypothesis test for correlations and partial correlations," *Psychonomic Bulletin & Review*, vol. 19, no. 6, pp. 1057–1064, 2012.

## Research Article

# Computer Vision-Driven Evaluation System for Assisted Decision-Making in Sports Training

Lijin Zhu 

*School of Physical Education and Health, Heze University, Heze 274015, China*

Correspondence should be addressed to Lijin Zhu; zhulijin@hezeu.edu.cn

Received 7 July 2021; Accepted 17 August 2021; Published 26 August 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Lijin Zhu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Computer vision has become a fast-developing technology in the field of artificial intelligence, and its application fields are also expanding, thanks to the rapid development of deep learning. It will be of great practical value if it is combined with sports. When a traditional exercise assistance system is introduced into sports training, the athlete's training information can be obtained by monitoring the exercise process through sensors and other equipment, which can assist the athlete in retrospectively analyzing the technical actions. However, the traditional system must be equipped with multiple sensor devices, and the exercise information provided must be accurate. This paper proposes a motion assistance evaluation system based on deep learning algorithms for human posture recognition. The system is divided into three sections: a standard motion database, auxiliary instruction, and overall evaluation. The standard motion database can be customized by the system user, and the auxiliary teaching system can be integrated. The user's actions are compared to the standard actions and intuitively displayed to the trainers as data. The system's overall evaluation component can recognize and display video files, giving trainers an intelligent training platform. Simulator tests are also available. It also demonstrates the efficacy of the algorithm used in this paper.

## 1. Introduction

With the rapid growth of China's emerging sports [1–3] and health industry [4–6], more people are devoting their time to sports like golf and skiing. Beginners who do not learn in a systematic way are more likely to fail to improve their technical level due to nonstandard movements, which can lead to sports injuries [7–9]. As a result, you must frequently review and analyze your actions in order to make improvements. Traditional training methods require professional sports coaches to conduct one-to-one teaching [10]. High labor costs and a lack of flexibility are among the issues. It is currently an urgent problem to solve how to conduct teaching and training in a simpler and more effective manner. Sports assist systems are frequently used by professional athletes in their daily training. Professional training analysts model the athletes based on sensor data and correct the details of the athletes' movements by analyzing the data collected from the limbs during training.

In recent years, with the emergence of big data and the substantial increase in computer parallel computing capabilities, deep learning [11–14] has made breakthroughs in the fields of computer vision [15–17] and natural language processing by relying on rich training data and powerful feature expression capabilities [18]; especially in the goals, many practical breakthroughs have been made in detection, machine translation [19], and motion recognition. Target detection is an important research topic of computer vision. It is the basis of many computer vision tasks [20, 21]. It has also been widely used in real life. For example, traditional target detection algorithms such as face recognition [22–24], unmanned driving, and target tracking use artificial manually extract features; there are problems such as incomplete feature information extraction and poor recognition effect. With the advent of convolutional neural networks, deep convolutional neural networks can automatically extract task-related feature information from massive data. Compared with traditional machine learning algorithms

[25–27], they have obvious advantages by manually setting feature extraction rules [28]. The detection algorithm has been greatly improved in detection accuracy and speed based on the goal of deep learning, which not only broadens the scope of computer vision applications but also adds more application value. Another important research topic in computer vision is semantic analysis of video, particularly human action recognition, which has a wide range of application scenarios, including human-computer interaction and gesture recognition. It is primarily intended for video sequences. The video can no longer be accurately described by the characteristics of pure static images. Therefore, the human motion recognition algorithm [29, 30] not only needs to extract the spatial features of each frame of image but also analyzes the temporal features between frames. The development of recurrent neural networks makes deep learning capable of processing sequence data and is widely used in natural language processing tasks such as machine translation [31, 32], text generation, and personalized recommendation [33].

This article discusses combining computer vision technology and deep learning technology with sports and applying them to sports training and evaluation, providing a data-based training platform for sports trainers.

The main contributions of this paper are as follows:

- (1) This paper proposes a sports assisted evaluation system that uses deep learning algorithms to recognize human posture and provide an intelligent training platform for sports trainers
- (2) The concept of “close Angle” is proposed in this paper to solve the motion recognition error caused by inconsistencies in limb lengths and angles between the human body and the camera acquisition equipment. The cosine angle of known three-point coordinates can be used to calculate the joint angle of key parts of the human body during movement
- (3) To improve the model’s performance, this paper uses the idea of finding the Euclidean distance between two vectors to calculate the similarity between actions

The organization of the paper is as given. Section 2 discusses the background to the proposed research. Methodology of the paper is given in Section 3 with details of the work done in the proposed research. Experiments and results are given in Section 4. The paper is concluded in Section 5.

## 2. Background

Types of athletic, entertainment, mass, and medical movement include: type of competitive sports is people want to beat the opponent, and obtain good results, the level greatly cultivate and strengthen itself, groups in the body, such as physical, emotional bias will carry out the objective and reasonable, a full range of training and contest; type of competitive sports is people want to beat the opponent, and obtain good results, the level greatly cultivate and strengthen itself,

groups in the body, such as physical, emotional bias will carry out the objective and reasonable. Recreational sports are activities that people engage in during their free time or at designated locations in order to achieve a happy effect. This kind of activity has the characteristics of nonprofessional, leisure, and pleasure. The activities usually include ball games, chess and cards, travel and play, and ethnic activities of partial distribution. Mass sports are sports activities commonly held in real life, such as people want to improve their physical fitness, resist diseases, train candidates for elite, and leisure time. The main objects involved are workers, farmers, and regional groups. At the same time, different projects have been held for people with different genders and inconvenient movement. Medical sports are in the treatment of some diseases and injuries, with the help of sports skills to help diagnose and to help improve the body function.

The movement of the human body can be described by the movement of a few key parts, and most of the movements can be described by piecing and tracking the joint parts. Generally speaking, traditional sports mainly have four major disadvantages: the restriction of places and devices, the special teaching of difficult movements, the difficulty of recording the practice data, and the boring process of practice. As a result, the development of an auxiliary training system allows trainers to undertake sports training at any time and in any area, regardless of time or location, and to capture real-time training data. However, in recent years, researchers have begun to look for ways to use machines to analyze athletes’ movements. Cameras can capture the entire training process and use computer vision technology to analyze and process the captured image data, resulting in a machine recognition effect of the human body, which aids human research in motion to some extent.

## 3. Methodology

*3.1. Human Body Gesture Recognition.* We use OpenPose [34], an open source library for human posture recognition, as the research tool in this article. OpenPose uses VGGNet-19 as a feature extractor during network model training. One of the biggest features of VGGNet is that it can process input image data of different resolutions, so in system development. In the process, proper processing of the image data input to the network can improve the fluency of the system. OpenPose is a multiperson gesture recognition library developed according to a bottom-up approach. The bottom-up method is more robust in the early stage. No matter the effect is as good in single-person and multiperson gesture recognition, the developed system can be easily turn to multiplayer sports evaluation. The current mainstream human pose recognition libraries include OpenPose, Mask R-CNN, and Alpha-Pose. The human body recognition time is compared among the three, and it is found that as the number of people in the image increases, the running time of Mask R-CNN and Alpha-Pose shows linear growth, and the running time of OpenPose remains unchanged, so the system developed using OpenPose will be more stable.

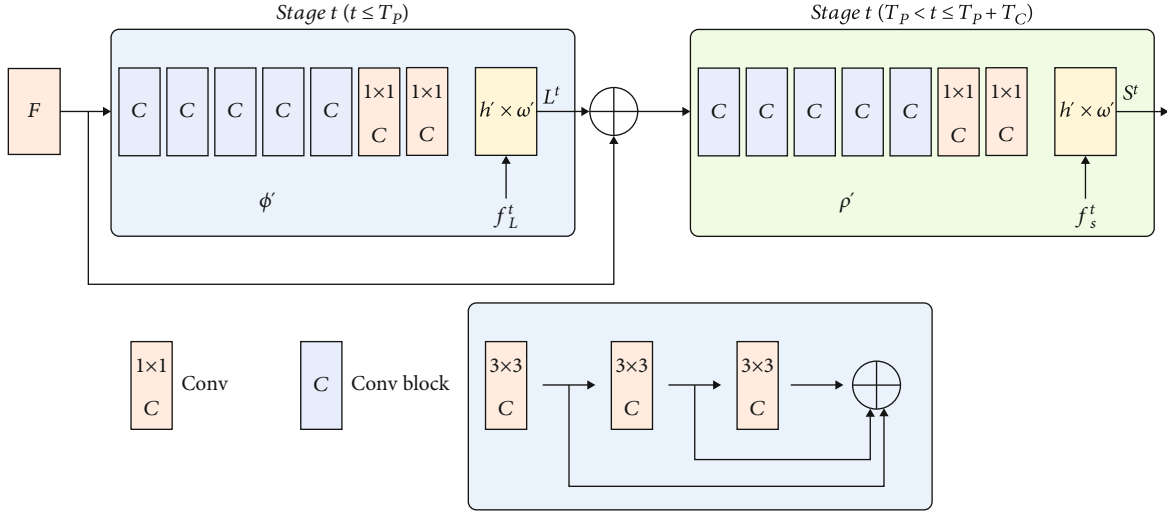


FIGURE 1: Schematic diagram of OpenPose network structure.



FIGURE 2: Example of key part connection.

The human body gesture recognition module's purpose is to choose a technology that will allow you to quickly and easily extract the position of key points on the human body from an image or video. For subsequent follow-up, the position information of the key points extracted must be accurate and conform to the actual rules of motion. The action evaluation went off without a hitch. After the action collection is complete, the human body gesture recognition module analyzes the image or video to extract the gesture data. Figure 1 depicts the OpenPose network structure.

In the first stage, the network generates a set of partial affinity fields  $L^1 = \phi^1(F)$ , and  $\phi^1$  represents the convolutional neural network structure that is also predicted in stage 1. In each subsequent stage, the prediction results from the previous stage and the original map feature  $F$  are cascaded and used to generate accurate predictions.

$$L^t = \phi(F, L^{t-1}), \forall 2 \leq t \leq T_p. \quad (1)$$

Here,  $\phi^t$  refers to the convolutional neural network structure of the prediction stage  $t$ , and  $T_p$  refers to the total

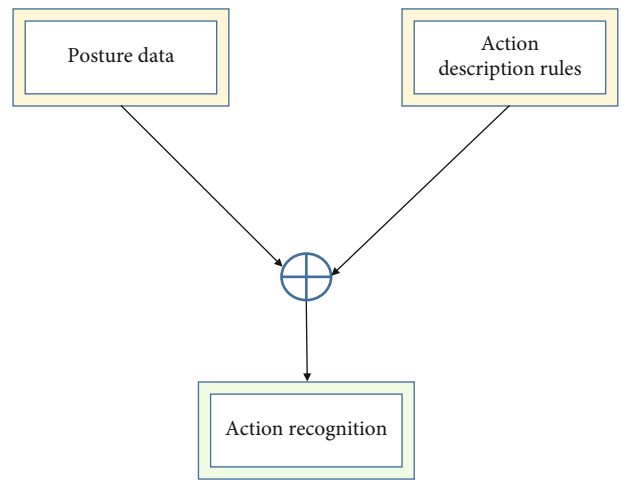


FIGURE 3: Schematic diagram of action description rules.

PAF prediction. After  $T_p$  iterations, the process is repeated to predict the confidence map from the latest PAF prediction stage.

$$\begin{aligned} S^{T_p} &= \rho^t(F, L^{T_p}), \forall t = T_p, \\ S^t &= \rho^t(F, L^{T_p}, S^{t-1}), \forall T_p < t \leq T_p + T_C. \end{aligned} \quad (2)$$

Here,  $\rho^t$  refers to the convolutional neural network structure of the prediction stage  $t$ , and  $T_C$  refers to the total confidence map prediction stage. The component association strategy is shown in Figure 2.

Figure 2(a) shows the two body parts (neck and marrow) and all possible joint pairs of three people. Figure 2(b) shows the result of using the midpoint strategy to connect. Both the correct connection and the wrong connection meet the constraints.

**3.2. Action Evaluation.** The prerequisite for recognizing human actions in images or videos is that there must be a set of action description rules. The action recognition effect is achieved by processing the posture information data extracted by the human posture recognition module according to the action description rules. The workflow of action description rules in the action evaluation process is shown in Figure 3.

**3.2.1. Joint Angle.** The joint points of the human body all have a coordinate position in the image, and the method of calculating the cosine angle by knowing the three-point coordinates can completely calculate the joint angle of the human body correctly. The joint angle is introduced to describe the action. The calculation process does not need to know the length of the body's limbs, and the two different actions can be compared, and then, the experimental verification is carried out. The human skeleton model and bone label are shown in Figure 4.

The three-point coordinate is known to be based on the idea of finding the joint angle between two vectors using the law of cosines, as shown in Figure 5.

The calculation equation for finding the joint angle is as follows:

$$\begin{aligned} \overrightarrow{AB} &= (x_2 - x_1, y_2 - y_1), \\ |AB| &= ((x_2 - x_1)^2 + (y_2 - y_1)^2)^{1/2}, \\ \overrightarrow{BC} &= (x_3 - x_2, y_3 - y_2), \\ |BC| &= ((x_3 - x_2)^2 + (y_3 - y_2)^2)^{1/2}. \end{aligned} \quad (3)$$

Then, we can find the cosine of the joint angle  $B$ :

$$\cos \angle B = \frac{\overrightarrow{AB} \times \overrightarrow{BC}}{|AB||AC|}. \quad (4)$$

**3.2.2. Action Similarity.** If there are two different athletes doing the same movement and you want to know who is doing more standard, you need to use a method to find

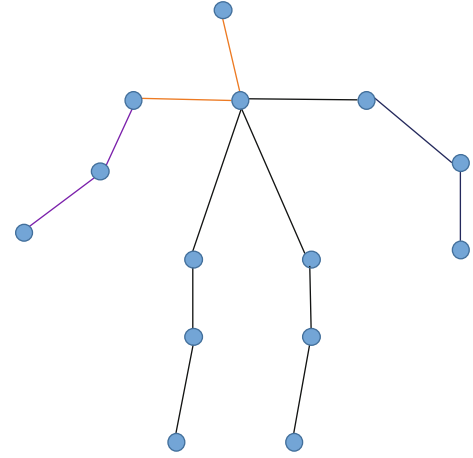


FIGURE 4: Schematic diagram of human skeleton.

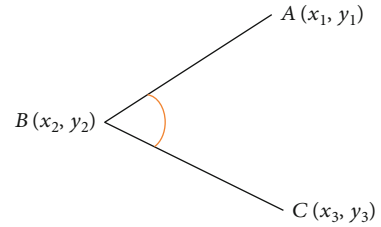


FIGURE 5: Schematic diagram of human skeleton.

out the “distance” between the two movements. Here, the concept of movement similarity is introduced. Action similarity is at the heart of action description rules; it can be used to identify a single action, and it can also be used to identify the beneficial actions in continuous action sequences based on single action recognition. Using the joint angle data of the two movements, we can find the distance between the two movements, and this distance is actually the degree of similarity between the movements. The joint angle data of the two movements were, respectively, expressed in the form of multidimensional vectors, such as  $A_1, A_2, \dots, A_n$  and  $T_1, T_2, \dots, T_n$ , where  $A$  represents the joint angle of the action to be measured and  $T$  represents the joint angle of the template action. The Euclidean distance was used to solve the distance between the two vectors:

$$d(x, y) = \left( \sum_{i=1}^n (x_i - y_i)^2 \right)^{1/2}. \quad (5)$$

Then:

$$\text{distance} = ((T_1 - A_1)^2 + (T_2 - A_2)^2 + (T_3 - A_3)^2 + \dots + (T_n - A_n)^2)^{1/2}. \quad (6)$$

The smaller the distance, the more similar the action.

**3.2.3. Action Evaluation.** We use the minimum Euclidean distance, which is the similarity measure between two

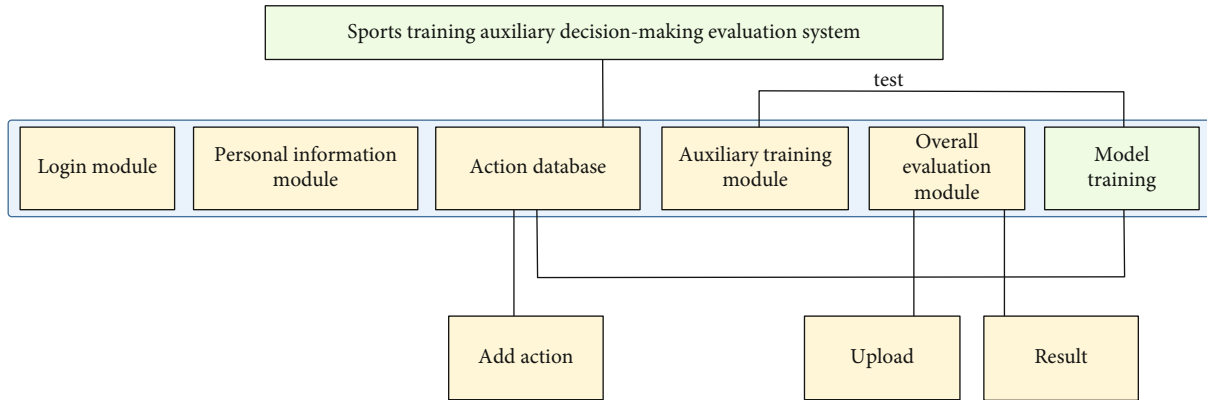


FIGURE 6: Schematic diagram of the overall model.

actions. The system can allow each joint angle between the movement of the exercise trainer and the standard movement to have the same error range. Within this range standard, find the movement most similar to the standard movement in the continuous movement. The specific realization can be that the realization prepares  $m$  standard action data and then finds the  $x$  (where  $x \leq m$ ) key actions that are most similar to the standard action in a motion video containing  $n$  (where  $n > m$ ) actions.

First, select a representative key action in the movement as the standard action, and prepare a picture that contains the standard action. Use the human body gesture recognition technology to extract the joint points of the characters in the picture, calculate the respective joint angles, and press the action. Sort them one after the other (such as action 1 and action 2) and save them in the database. Then, use the shooting equipment to record the complete video of the athlete making this movement and import it into the system. The system extracts the joint angle data of all the movements in the entire video. Finally, the system compares the test motions' joint angle data to the standard motion database, finds the test motions that are the most similar to the standard motions and have the same number, and displays the difference between the test motions and the standard motions to the user. The user has the option of displaying according to the view. The difference identifies the source of the tester's issue.

**3.3. Sports Training Auxiliary Decision-Making Evaluation System.** The exercise-assisted evaluation system we proposed includes system login entry, account information modification, standard action database collection, assisted teaching, and overall exercise evaluation. The login window requires the user to enter the correct user name and password to enter the system. The account information allows the user to enter the system. Change the account password in time when the account is abnormal. The standard action database allows users to set a key action during exercise that can be used as the evaluation standard. The auxiliary teaching allows users to practice exercises anytime and anywhere. The overall evaluation can be based on the user's prerecorded exercises. The video performs action filtering and outputs all action data that are highly similar to standard

TABLE 1: Action similarity evaluation result.

Sports actions	Euclidean distance
1	5.236
2	4.958
3	3.658
4	5.011
5	4.125
6	3.256

actions. The overall functional module diagram is shown in Figure 6.

## 4. Experiments and Results

**4.1. Experimental Setup.** All of the experiments in this article were run on a deep learning server with two NVIDIA GTX 1080 TI graphics cards with 11 GB of memory. The system software platform is primarily developed in the Visual Studio 2015 environment. The open source computer vision library OpenCV for deep neural networks, Microsoft's MFC interface library, CUDA architecture, and GPU acceleration library CUDNN were all used in the development process. MFC is a set of basic class libraries created by Microsoft using the C++ programming language.

**4.2. Dataset.** This paper uses GolfDB as the experimental data set, which is a high-quality golf swing video data set, which is specially created for golf swing motion and used for golf swing motion recognition. GolfDB contains 1,400 golf swing video samples, with a total of more than 390 K frames of video data, which are collected manually from YouTube video website. 580 regular-speed and slow-motion golf swing motion videos are the most important for golf swing motion recognition. To ensure that the club is visible and to reduce motion blur, the sampling only considers 30 fps, 720P resolution video. The video mainly intercepts the swing movements of 248 professional golfers from the game videos of PGA, LPGA, and Champions Tours.



**4.3. Analysis of Auxiliary Training Evaluation Results.** The auxiliary teaching function is actually the concrete application of the evaluation method of a single movement. Users can select the corresponding actions through the drop-down list box for individual learning and click the detection button to start training. The computer will enable its own camera to capture the human body in real time and display it in the picture control. At the same time, the body posture information and joint angle data of each frame will be output in real time on the right side of the picture space. The eight edit boxes on the far right with read-only properties display the data of the current joint angle of the action in real time.

Taking the movement as the test target, the user makes the corresponding movement in front of the camera, and the system automatically identifies the joint angle data of the current posture and calculates the Euclidean distance from the selected standard movement. When the current distance value is less than one value, the training stops. The function of Euclidean theorem in auxiliary teaching is to set a Euclidean distance threshold. When the Euclidean distance between the trainer's action and the joint angle vector of the standard action is less than a specific value, the training will be stopped in time, and the training results of the current action will be output. The training results include the detailed display between the eight joint angles. After stopping the exercise, the system quickly displays the joint angle data of the last frame and the standard movement in the list control in the lower right corner, so that the trainer can observe the difference between himself and the standard movement and make self-adjustment.

It can be seen from Table 1 that the Euclidean distance value of action 3 and action 6 is the smallest, which proves that the exercise is in compliance with the standard. At the same time, the Euclidean distance of action 1 and action 4 is larger, and the sports training decision-making department should focus on action 1, and action 4 conducts intensive training and correction.

In the actual video motion recording, the speed of the movement process of the demonstration personnel will not have any influence on the experimental results, because the system will pick out all the motion frames that meet the conditions and list the corresponding joint angle data. The user can select all the optimal actions, that is, the actions with the minimum distance value, as the basis of the whole set of actions. With the help of the motion assistant evaluation system based on human posture recognition developed in this paper, the included angle at the joint can be obtained, and the distance between movements can be calculated according to the Euclidean theorem. After the comparative analysis of the experiment, it has been proved that it can be applied to the auxiliary teaching and the auxiliary evaluation of sports. Through the comparison data of joint angle given by this system, athletes can find their own problems by checking the difference between their movements and standard movements and make improvements according to the problems. With more practice, they will get closer and closer to the standard movements.

## 5. Conclusion

In this article, we propose a motion assistance evaluation system based on human posture recognition based on deep learning algorithms. The system is mainly composed of three parts: standard motion database, auxiliary teaching, and overall evaluation. System users can customize the standard motion database to assist teaching. Some systems can compare the user's actions with standard actions and visually display them to the trainer in the form of data. The overall evaluation system can identify and filter video files to provide an intelligent training platform for trainers. In addition, simulation experiments also prove the effectiveness of the algorithm in this paper.

## Data Availability

The data used to support the findings of this study are included within the article.

## Conflicts of Interest

All the authors do not have any possible conflicts of interest.

## References

- [1] T. Meiklejohn, D. L. Ferkins, and I. O'Boyle, *Exploring Governance Design for New and Emerging Sports: The Case of Stand up Paddling in New Zealand*, Unitec Institute of Technology, New Zealand, 2017.
- [2] E. R. McCollum, *The Impact of the NCAA Emerging Sports Program for Women on Title IX Compliance: A 10 Year Longitudinal Study*, [Ph.D. thesis], The University of North Carolina at Chapel Hill, New Zealand, 2017.
- [3] V. Mitrousias, G. Halatsis, I. Bampis, A. Koutalos, G. Psareas, and A. Sakkas, "Epidemiology of injuries in pole sports: emerging challenges in a new trend," *British Journal of Sports Medicine*, vol. 51, no. 4, 2017.
- [4] A. B. Cohen, E. R. Dorsey, S. C. Mathews, D. W. Bates, and K. Safavi, "A digital health industry cohort across the health continuum," *NPJ digital medicine*, vol. 3, no. 1, 2020.
- [5] M. Mackintosh, J. Mugwagwa, G. Banda et al., "Health-industry linkages for local health: reframing policies for African health system strengthening," *Health Policy and Planning*, vol. 33, no. 4, pp. 602–610, 2018.
- [6] A. Zuniawan, S. M. Wirawati, M. Fahlevi, A. Purwanto, N. A. Vizano, and R. Pramono, "Did Seiri Seiton Seiso Seiketsu and Shitsuke affected medical health industry business performance?," *European Journal of Molecular & Clinical Medicine*, vol. 7, no. 7, pp. 97–114, 2020.
- [7] C. V. Andreoli, B. C. Chiamonti, E. Biruel, A. C. Pochini, B. Ejnisman, and M. Cohen, "Epidemiology of sports injuries in basketball: integrative systematic review," *BMJ Open Sport & Exercise Medicine*, vol. 4, no. 1, article e000468, 2018.
- [8] R. Pol, R. Hristovski, D. Medina, and N. Balague, "From microscopic to macroscopic sports injuries. Applying the complex dynamic systems approach to sports medicine: a narrative review," *British Journal of Sports Medicine*, vol. 53, no. 19, pp. 1214–1220, 2019.
- [9] A. Gledhill, D. Forsdyke, and E. Murray, "Psychological interventions used to reduce sports injuries: a systematic review of

- real-world effectiveness,” *British Journal of Sports Medicine*, vol. 52, no. 15, pp. 967–971, 2018.
- [10] A. Hershkovitz and O. Karni, “Borders of change: a holistic exploration of teaching in one-to-one computing programs,” *Computers & Education*, vol. 125, pp. 429–443, 2018.
- [11] Y. Tong, L. Yu, S. Li, J. Liu, H. Qin, and W. Li, “Polynomial fitting algorithm based on neural network,” *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 1, pp. 32–39, 2021.
- [12] X. Ning, Y. Wang, W. Tian, L. Liu, and W. Cai, “A biomimetic covering learning method based on principle of homology continuity,” *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 1, pp. 9–16, 2021.
- [13] R. Liu, X. Ning, W. Cai, and G. Li, “Multiscale dense cross-attention mechanism with covariance pooling for hyperspectral image scene classification,” *Mobile Information Systems*, vol. 2021, Article ID 9962057, 15 pages, 2021.
- [14] J. Zhang, Y. Liu, H. Liu, and J. Wang, “Learning local–global multiple correlation filters for robust visual tracking with Kalman filter redetection,” *Sensors*, vol. 21, no. 4, p. 1129, 2021.
- [15] J. Zhang, J. Sun, J. Wang, and X. G. Yue, “Visual object tracking based on residual network and cascaded correlation filters,” *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, 2021.
- [16] C. Yan, G. Pang, X. Bai et al., “Beyond triplet loss: person re-identification with fine-grained difference-aware pairwise loss,” *IEEE Transactions on Multimedia*, p. 1, 2021.
- [17] J. Zhang, X. Jin, J. Sun, J. Wang, and A. K. Sangaiah, “Spatial and semantic convolutional features for robust visual object tracking,” *Multimedia Tools and Applications*, vol. 79, no. 21–22, pp. 15095–15115, 2020.
- [18] Y. Lu, W. Li, X. Ning et al., “Blind image quality assessment based on the multiscale and dual-domains features fusion,” *Concurrency and Computation: Practice and Experience*, no. - article e6177, 2021.
- [19] C. Ying, Y. Shuyu, L. Jing, D. Lin, and Q. Qi, “Errors of machine translation of terminology in the patent text from English into Chinese,” *ASP Transactions on Computers*, vol. 1, no. 1, pp. 12–17, 2021.
- [20] X. Zhang, Y. Yang, Z. Li, X. Ning, Y. Qin, and W. Cai, “An improved encoder-decoder network based on strip pool method applied to segmentation of farmland vacancy field,” *Entropy*, vol. 23, no. 4, p. 435, 2021.
- [21] W. Cai, Z. Wei, R. Liu, Y. Zhuang, Y. Wang, and X. Ning, “Remote sensing image recognition based on multi-attention residual fusion networks,” *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 1, pp. 1–8, 2021.
- [22] X. Ning, P. Duan, W. Li, and S. Zhang, “Real-time 3D face alignment using an encoder-decoder network with an efficient deconvolution layer,” *IEEE Signal Processing Letters*, vol. 27, pp. 1944–1948, 2020.
- [23] S. Li, X. Ning, L. Yu et al., “Multi-angle head pose classification when wearing the mask for face recognition under the COVID-19 coronavirus epidemic,” in *2020 international conference on high performance big data and intelligent systems (HPBD&IS)*, pp. 1–5, Shenzhen, China, 2020.
- [24] L. Sun, W. Li, X. Ning, L. Zhang, X. Dong, and W. He, “Gradient-enhanced softmax for face recognition,” *IEICE Transactions on Information and Systems*, vol. E103.D, no. 5, pp. 1185–1189, 2020.
- [25] Y. Jiang, X. Gu, D. Wu et al., “A novel negative-transfer-resistant fuzzy clustering model with a shared cross-domain transfer latent space and its application to brain CT image segmentation,” *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 18, no. 1, pp. 40–52, 2021.
- [26] R. Liu, W. Cai, G. Li, X. Ning, and Y. Jiang, “Hybrid dilated convolution guided feature filtering and enhancement strategy for hyperspectral image classification,” *IEEE Geoscience and Remote Sensing Letters*, 2021.
- [27] M. Gao, W. Cai, and R. Liu, “AGTH-net: attention-based graph convolution-guided third-order hourglass network for sports video classification,” *Journal of Healthcare Engineering*, vol. 2021, Article ID 8517161, 10 pages, 2021.
- [28] M. Li, G. Zhou, W. Cai et al., “MRDA-MGFSNet: network based on a multi-rate dilated attention mechanism and multi-granularity feature sharer for image-based butterflies fine-grained classification,” *Symmetry*, vol. 13, no. 8, p. 1351, 2021.
- [29] H. Kuehne, H. Jhuang, E. Garrote, T. Poggio, and T. Serre, “HMDB: a large video database for human motion recognition,” in *2011 International Conference on Computer Vision*, pp. 2556–2563, Barcelona, Spain, 2011.
- [30] C. Sminchisescu, A. Kanaujia, and D. Metaxas, “Conditional models for contextual human motion recognition,” *Computer Vision and Image Understanding*, vol. 104, no. 2-3, pp. 210–220, 2006.
- [31] J. Xiao, Y. Dai, and X. Shi, “Translation and influence of one two three... infinity in China,” *ASP Transactions on Computers*, vol. 1, no. 1, pp. 18–23, 2021.
- [32] L. Liang, Q. Yin, and C. Shi, “Exploring proper names online and its application in English teaching in university,” *ASP Transactions on Computers*, vol. 1, no. 1, pp. 24–29, 2021.
- [33] W. Cai, Y. Song, and Z. Wei, “Multimodal data guided spatial feature fusion and grouping strategy for e-commerce commodity demand forecasting,” *Mobile Information Systems*, vol. 2021, Article ID 5568208, 14 pages, 2021.
- [34] Z. Cao, G. Hidalgo, T. Simon, S. E. Wei, and Y. Sheikh, “OpenPose: realtime multi-person 2D pose estimation using part affinity fields,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 1, pp. 172–186, 2019.

## Research Article

# Application of Flipped Classroom Model Driven by Big Data and Neural Network in Oral English Teaching

Yujun Zeng 

College of Foreign Languages of Hunan Institute of Engineering, Xiangtan 411101, China

Correspondence should be addressed to Yujun Zeng; 05012@hnie.edu.cn

Received 10 July 2021; Accepted 12 August 2021; Published 26 August 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Yujun Zeng. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the advancement of big data and neural network technology, flipped classroom informatization has shifted the traditional order of knowledge transfer and internalization, emphasizing students' autonomous learning before class, knowledge absorption, and knowledge completion in class with the assistance of teachers. Students' internalization and consolidation create the conditions for individualized learning. In foreign teaching, the benefits and feasibility of the flipped classroom have been demonstrated, and it is a promising new teaching model. Although recent research on oral English teaching in Chinese universities has yielded promising results, students' classroom activity and participation remain low, learning initiative is lacking, and opportunities and time for oral training are insufficient. This article uses flipped classroom, big data, and neural network technology to teach college oral English classes, with the goal of determining whether the flipped classroom model can help students improve their oral English proficiency and self-learning ability, as well as exploring students' attitudes toward the flipped classroom model. This paper first proposes a big data and deep neural network-based algorithm for detecting oral English pronunciation errors, which can be used for self-correction of students in the flipped classroom mode to improve the quality of oral English teaching. Finally, we also conducted simulation experiments, and the experimental results show that our algorithm is 4.12% better than SVM.

## 1. Introduction

Although a series of teaching reforms [1–3] have been carried out, there are still many problems in English teaching in China [4–6]. For example, students lack the ability of independent learning [7–9] and cooperative learning [10–12]. The teacher's lack of professional quality and the solidification of educational concept are the unfavorable factors that affect the teaching effect. "Flipped classroom" teaching mode [13–16] is a new form of teaching and classroom organization. Many new inspirations and concepts will be integrated and put into practice in the teaching process of flipped classroom. It not only stimulates students' interest in learning, but also develops students' independent learning ability and cooperative research ability. Flipped classroom subverts the traditional classroom model [17] and also changes the role of teachers and students in traditional teaching. The time in the classroom is rearranged, and the time between knowledge transfer in class

and internalization of knowledge absorption after class is adjusted, resulting in traditional classroom innovation.

Big data [18] and neural network technology [19–21] have transformed how people communicate. People are constantly learning new things and interacting with others. Technology has had a significant impact on not only how people work and live, but also on how they learn. Digital technology is used to record lecture content, as well as audio and video related to teaching knowledge, providing technical support for flipped classroom implementation. Furthermore, after class, students must self-study online knowledge content, freeing up valuable classroom time to address students' questions and difficulties or engage in other constructive classroom activities. Instead of doing what the teacher says in class and doing homework and exercises after class, you should do what the teacher says. The teaching method is flipped, with students receiving explicit instructions to learn after class and then applying what they have learned in class.

The most significant advantage of flipped classroom is that it can maximize the effect of classroom learning [22], which is also the most basic demand of teaching. In the liberated classroom time, the teacher can give personalized guidance to each learner according to his/her characteristics and knowledge mastery, so as to make the classroom time more meaningful. Moreover, more flexible learning tasks and learning pace can be provided in the schedule according to each student's learning progress. Educators have benefited from technological advancements in education for a long time by incorporating technology into their teaching. Students can access online learning at any time and from any location thanks to popular online services like online video sites and Baidu Cloud. Students can easily follow a variety of online courses by clicking on links on social media, which is very convenient and flexible. As a result, a lot of flipped classroom big data is generated.

Based on the foregoing observations, we can conclude that, while recent research on oral English teaching [23] in Chinese universities has yielded promising results, students' activity and participation in the classroom remain low, learning initiative is lacking, and opportunities and time for oral training are insufficient. This article explores students' attitudes toward flipped classroom teaching mode by applying flipped classroom, big data, and neural networks [24–26] to college oral English classroom teaching, which helps to improve students' oral English level and autonomous learning ability.

The main contributions of this paper are as follows:

- (1) This paper applies flipped classroom, big data, and neural network technology to college oral English classroom teaching, which helps to improve students' oral English level and autonomous learning ability, and explores students' attitudes toward flipped classroom teaching mode
- (2) This paper proposes a pronunciation error detection algorithm based on phonetics space and KL divergence distance metric. The basic element of the phonetics space is senone, and its posterior probability is obtained by discriminative learning from the acoustic features by the deep neural network

The rest of the paper is organized as follows. In Section 2, the background is outlined. The methodology is conducted in Section 3. The experimental results are further summarized in Section 4. Finally, Section 5 concludes the paper with a summary and future research directions.

## 2. Background

At present, emerging information technology [27, 28] provides a powerful driving force for economic and social change and development. Facing the knowledge explosion in the context of the new era and the rapid changes in technology, learners have new learning needs. In the information age, how to cultivate people also puts forward new requirements. Make full use of information technology and strive

to create an information-based learning environment for learners, so as to promote the reform of teaching concepts, teaching models, and teaching content. In the daily teaching process, information technology is continuously deepened and widely used, so as to better serve the new talent training services under the background of the times. In the context of the new era, the intelligentization of education should play an active leading role, and the intelligentization of education should support and lead the modernization of education. Simultaneously, education informatization is critical to the renewal of educational concepts, the reform of teaching models, and the reconstruction of the educational system. The incorporation of big data and neural network technology into the process of subject teaching innovation, as well as the realization of deep integration between information technology and subject teaching, has become a hot topic of current research in this context. Teaching concepts and teaching activity design continue to innovate and present a diverse development as a result of continuous advancement and in-depth research.

In terms of classroom form, teaching process, and teacher and student roles, the flipped classroom has been flipped. The format has shifted from “teaching first, then learning” to “first, then teaching.” The teaching process is being turned on its head thanks to information technology. Teachers are now “teaching,” and the traditional role of students as “learners” has shifted. Flipped classrooms overcome the limitations of traditional teaching. Teaching places a greater emphasis on students' learning processes, allowing them to have a more authentic learning experience. In the teaching process, more attention is paid to student self-study, teacher-student interaction, and teacher-student, student-student interaction. Collaborative exchanges between students have changed the former difficulty of passively accepting knowledge and help students develop their lifelong learning capabilities. From its emergence to its continuous development, flipped classrooms have been sought after by different scholars and educators all over the world and have had a profound impact. Objectively speaking, the flipped classroom has improved the teaching quality of the subject and the ability of students, and it has had an important reform effect on classroom teaching. However, flipped classrooms have not set off a wave of reform in China. Aside from the higher demands of flipped classrooms in a networked teaching environment, teachers' educational concepts and teaching viewpoints must be updated urgently. Nowadays, education is becoming increasingly informatized, and educational infrastructure is being built to a higher standard. To promote the depth between information technology and subject teaching, teachers should learn to use networked learning platforms to make full use of advanced teaching concepts such as blended learning and flipped classrooms.

## 3. Methodology

*3.1. Flipped Classroom.* Flipped classroom is a teaching mode that mixes direct explanation and constructivist learning. The traditional teaching model is mainly divided into three parts: preclass knowledge preview, in-class knowledge teaching, and after-class knowledge consolidation. The “flipped

classroom” breaks the traditional teaching model and reverses the process of knowledge transfer and internalization, that is, teachers send relevant learning materials to WeChat, QQ, and other online platforms before class, and students complete the teaching content independently before class. Classroom time is mainly used to carry out various teaching activities, such as teachers explaining the problems existing in students’ self-learning before class and students exploring problems in groups, so as to obtain a deeper level of new knowledge. Understand, digest, and absorb. Flipped classroom extends the classroom from school to home, greatly enhancing the openness of education. Students can formulate their own learning progress according to their own learning ability and will no longer be affected by the uniform teaching of teachers. This gives full play to the autonomy and initiative of students in learning, and individualized learning is also guaranteed, so as to achieve a better educational effect. The flipped classroom teaching model is shown in Figure 1.

### 3.2. Flipped Classroom Teaching Model of Spoken English

**3.2.1. Basic Model.** The timeline of a flipped classroom generally includes two parts: before class and during class. Before class, it is used for students to study independently, that is, the stage of “knowledge externalization.” The class is used for the development of learning activities, that is, the stage of “knowledge internalization.” Due to the differences in learners, implementation environment, learning methods, and other subjects in oral English teaching, the author takes Professor Robert Tallbert’s flipped classroom implementation structure model as the basis and combines the teaching content and teaching objects of this flipped classroom. Characteristic, the structure model of the flipped classroom implementation of Professor Robert Tallbert has been modified and refined accordingly, and the teaching implementation model of the flipped classroom of spoken English has been constructed. The specific model is shown in Figure 2.

The textualization of the flipped classroom teaching implementation model in Figure 2 can be divided into the following four steps in detail:

- (1) *Teachers Provide Learning Materials.* There are many types of preclass learning materials for flipped classrooms, such as teaching microvideos, presentations, and teaching cards. Teachers can choose according to teaching objects, teaching content, and actual teaching needs. But generally speaking, the preclass learning materials of the flipped classroom are mainly teaching microvideos, and teachers can download or make their own teaching resources from related websites. Famous foreign teaching video websites include Khan Academy and TED-ED. Teachers can look for video resources that match their teaching content among high-quality open educational resources as the course teaching content. After the teachers prepare the learning materials, they can upload them to the learning platform.
- (2) *Students’ Autonomous Learning Stage before Class.* First, students learn the learning materials provided by the teacher before class to complete the task of autonomous learning. If students encounter problems in the learning process, they can communicate and discuss with their classmates on the platform to find solutions. Problems that cannot be solved can be fed back to the teacher individually, and the teacher will give a one-to-one personalized explanation on the platform. Secondly, students independently complete the test questions uploaded by the teacher to the platform, and the teacher roughly determines the student’s preclass learning effect based on the students’ test situation. Finally, students will perform preclass exercises according to the oral tasks assigned in the self-learning task list, paving the way for oral English communication in the classroom.
- (3) *In-Class Knowledge Internalization Stage.* Before class, students have already learned the relevant teaching knowledge points of this class. This provides ample time for the internalization of knowledge in the middle of the class. First, the teacher will give a unified explanation of the questions with a higher error rate according to the students’ test situation before class and explain the relevant knowledge points from the wrong questions again. Second, conduct classroom testing. The methods of classroom testing can be diversified, such as teachers asking questions, students answering, or students asking and answering each other in a game. Third, let students discuss the knowledge expansion questions in the learning task list in groups and share the results of the discussion after the end. The teacher comments and supplements the results of the students’ discussion. Then, carry out classroom activities. As it is an oral class, teachers can set up some classroom activities aimed at improving students’ oral communication skills. For example, students are grouped to carry out activities such as situational dialogue, role playing, and microvideo dubbing. In the course of the activity, teachers should give individual tutoring to individual student with poor oral English, encourage them to speak English, actively participate in classroom activities, and encourage students to take the initiative to show their results after the end.
- (4) *Evaluation and Summary Stage.* Let students conduct self-evaluation and mutual evaluation on their own learning attitude, self-confidence, cooperative spirit, and spoken English expression in this lesson, cooperate in groups to summarize the key and difficult points of this lesson, and complete the unit review mind map. After the end, the group representatives went to the stage to show the unit mind map and explained the important and difficult points of this lesson, the teacher made comments and supplements, and the students made corresponding notes.

**3.3. Spoken Pronunciation Detection Algorithm.** Figure 3 depicts the structure of the proposed pronunciation error detection model in this paper. It is divided into two sections:

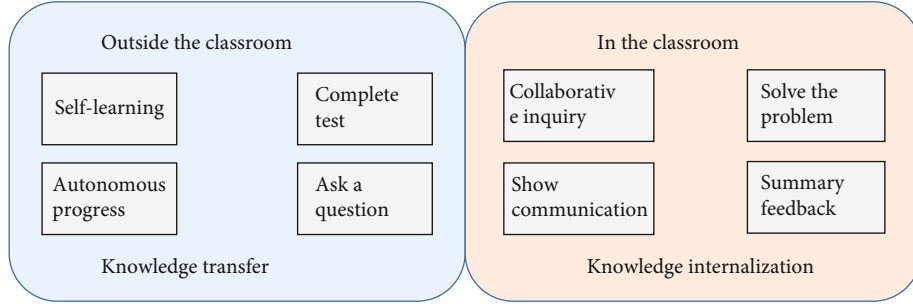


FIGURE 1: The teaching mode of flipped classroom.

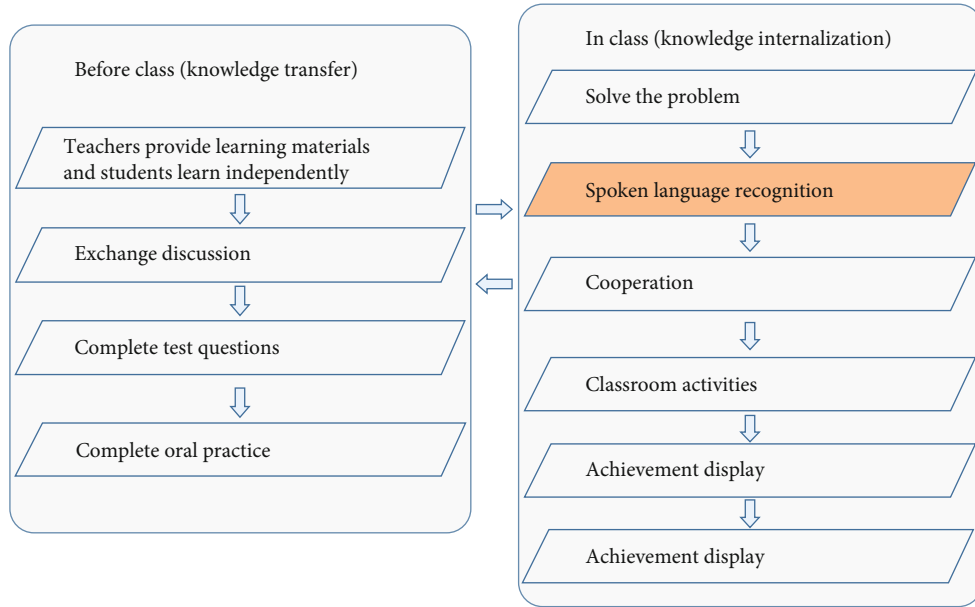


FIGURE 2: An implementation model of flipped classroom teaching of spoken English.

input characteristics and classifier design. The two parts will be discussed in greater depth below.

**3.3.1. Input Features.** In view of the fact that acoustic features, such as MFCC and PLP, are susceptible to noise interference from different speakers, transmission channels, and recording environment, this paper adopts high-dimensional phonetics features as the basic features of pronunciation error detection. Suppose there are  $M$  different senses  $\{s_1, s_2, \dots, s_M\}$  in the acoustic model, for any speech frame  $x_t$ , its characteristics in phonetics space are expressed as follows:

$$f(x_t) = [p(s_1|x_t), \dots, p(s_M|x_t)]^T, \quad (1)$$

where the posterior rate  $p(s_i|x_t)$  is the output value of the  $i$ th node of DNN.

Suppose that the phoneme of speech segment  $X$  is  $b$ , and three hidden states  $bs_1, bs_2, bs_3$  are obtained through DNN. For the speech segment corresponding to each state, the center vector is calculated to indicate its position in the speech space. The calculation equation is as follows:

$$\bar{f}(X) = [\bar{p}(s_1|X), \dots, \bar{p}(s_i|X), \dots, \bar{p}(s_M|X)]^T, \quad (2)$$

$$\bar{p}(s_i|X) = \frac{1}{t_{ek} - t_{sk} + 1} \sum_{t=1}^{t_{ek}} p(s_i|X_t). \quad (3)$$

If the total number of sensors in the acoustic model is  $M$ , and each contains 3 valid states, the dimension of the feature vector is  $3M$ . Feature direction  $f(X)$  will be used as the input feature of the phoneme classifier for classifier training.

**3.3.2. Classifier.** The task of detecting pronunciation errors for each phoneme is abstracted as a binary classification problem. Assuming that there are  $M$  different phonemes in the phoneme set, the traditional classification method divides the entire data set into  $M$  parts according to the corresponding phonemes and learns a binary classifier for each phoneme separately.

Assume that the training data set is  $S = \{X_n, t_n\}_{n=1}^N$ , which contains  $M$  different phonemes  $\{p_1, p_2, \dots, p_M\}$ , the phoneme label corresponding to the  $n$ th sample is  $C_n \in \{1, 2, \dots, M\}$ , and its pronunciation is correct or wrong is  $t_n \in \{0, 1\}$ , where 1 means correct pronunciation and 0 means wrong

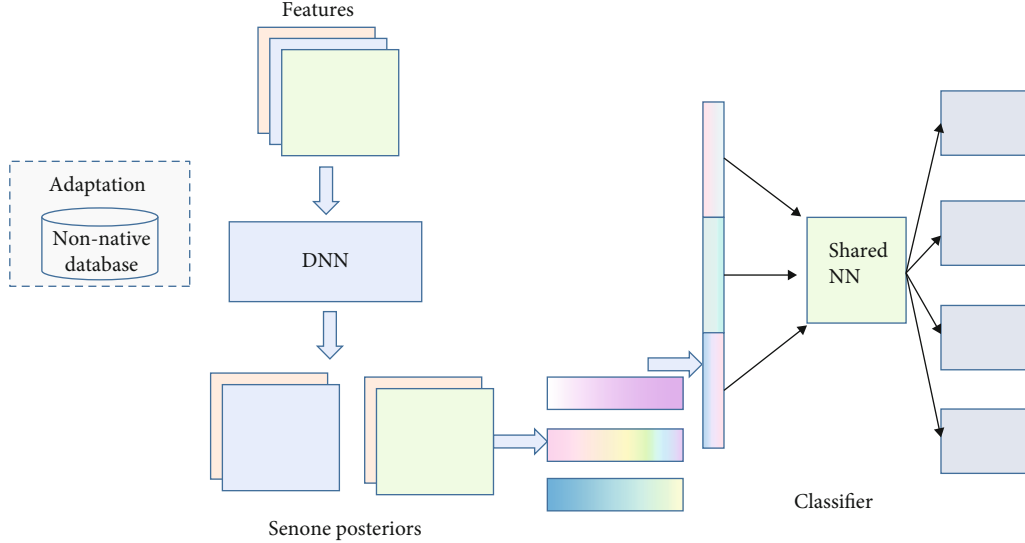


FIGURE 3: Pronunciation error detection model based on big data and deep neural network.

pronunciation. The maximum likelihood criterion is used to estimate the model parameters, and the optimization objective function is defined as follows:

$$\text{Loss} = \min(-\log L),$$

$$L = \prod_{n=1}^N \sum_{m=1}^M \delta(C_n = m) y_m(X_n)^{t_n} (1 - y_m(X_n))^{1-t_n}, \quad (4)$$

$$y_m(X_n) = \sigma_m(W_m^T \phi(X_n)).$$

In the testing phase, for sample  $X$ , if its corresponding phoneme is marked as  $i$ , then its category is judged as follows:

$$p(C|\phi(X)) = \sigma_i(W_i^T \phi(X)) \begin{cases} \geq \theta & \text{true} \\ < \theta & \text{false} \end{cases}, \quad (5)$$

where  $\theta$  is the system threshold.

## 4. Experiments and Results

**4.1. Experimental Environment and Parameters.** The programming language used is Python, the version is 3.6.5, the deep learning framework used is Keras2.1.5, and the IDE for program deployment is Pycharm, and all experiments are conducted in the same environment. All our experiments have been conducted on a desktop PC with an Intel Core i9-9900K processor and an NVIDIA GeForce GTX 2080ti GPU. We have implemented the model construction through the Keras deep learning library, the programming language we use is Python, and we batch processed 256 samples each time.

**4.2. Data Sets.** The standard pronunciation database is an English isolated word recognition database named LDC95S27, in which each sentence contains only one English word. The entire data set contains 93,667 sentences, with a total duration of 23 hours. We divide it into two parts: training set and test set. Among them, the training set contains 900 speakers,

TABLE 1: The distribution of the number of various pronunciation errors in the English word learning database.

Type	True	Substitution	Deletion	Insertion
Count	103522	15673	3180	1255

TABLE 2: The accuracy of pronunciation error recognition of different structure classifiers on the English test set.

Method	200 nodes	400 nodes	600 nodes	1200 nodes
1 layer	93.1	93.4	93.9	92.1
2 layers	93.2	92.9	93.3	92.2
3 layers	91.3	91.9	93.0	92.9

6,700+ words in different texts, and the total duration is 20 hours: the test set contains 80 speakers, and the total duration is about 3 hours. The English learning database is collected and recorded by the author. The database contains 60 Chinese people, each of whom reads 300 words, and the word text is randomly generated from LDC95S27. The phoneme-level misreading, interruption, and omission errors of each word are marked by a phonetic expert. The distribution of various pronunciation errors on the entire data set is shown in Table 1. It can be seen from the distribution table that misreading errors account for 77.9% of all pronunciation errors, which is the main type of error.

**4.3. Evaluation Index.** We scored by calculating the distance  $D_1$  and  $D_2$  between the correct pronunciation model  $C$  and the wrong pronunciation model  $IC$  of the sample center  $X$  and the phoneme  $b$ . The final pronunciation accuracy score at the phoneme level is expressed as follows:

$$\text{Score} = \frac{e^{-D_1}}{e^{-D_1} + e^{-D_2}}, \quad (6)$$

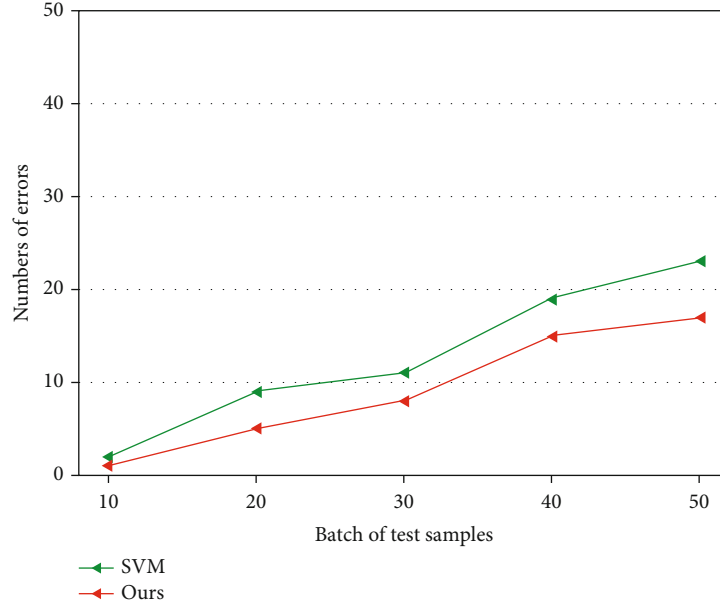


FIGURE 4: Confusion matrix.

$$\text{Score} \geq \theta \begin{cases} T & \text{true} \\ F & \text{false} \end{cases}. \quad (7)$$

**4.4. Experimental Results.** Firstly, the effects of different hidden layer network structures on experimental results are compared. Because the English learning database is relatively small, the underlying network structure used in the W experiment is also relatively small. The dropout algorithm is used in the training process to prevent overfitting. Table 2 shows the accuracy of pronunciation error detection on the verification set when the number of hidden layers is 1 layer, 2 layers, and 3 layers and the number of nodes in each layer is 200, 400, 600, and 1200. The system threshold is set to 0.4. As can be seen from the table, the artful response of different underlying network structures to the result of articulatory error detection is very small, and the change of accuracy is less than 1%. Among them, the result of the single hidden layer classifier is slightly better than the result of the multihidden layer classifier, which we believe may be caused by the relative shortage of training data. In view of this, the underlying network in subsequent experiments was all set as a single hidden layer, with a number of 600 nodes.

**4.5. Comparative Experiment.** Support Vector Machine (SVM) is usually used as a classifier in error detection algorithms based on dichotomy. In this section, we compare the performance of SVM and the model presented in this paper in terms of pronunciation error recognition on English learning databases. There are 40 phonemes in the English learning database, so 40 SVM classifiers need to be trained. All SVMs are trained with the SVM-Light tool, and the default parameters are set directly. For each phoneme in a sentence, the classifier predicts the conditional probability that it is pronounced correctly. By adjusting the system threshold value, the number of error recognition under different operation points can be

obtained. The pronunciation error recognition results of the two methods on the English learning test set are shown in Figure 4. It can be seen that the model in this paper is overall better than the SVM method, which proves the effectiveness of our model.

## 5. Conclusion

This paper uses flipped classroom, big data, and neural network technology to teach college oral English classes, with the goal of determining whether the flipped classroom model is effective in improving students' oral English proficiency in Chinese college oral English classes. Students' attitudes toward the flipped classroom teaching model are investigated using oral English ability and self-study ability. This article first proposes an algorithm for detecting oral English pronunciation errors based on big data and deep neural network technology, which can be used to help students self-correct in a flipped classroom setting and improve the quality of oral English instruction. Finally, we ran simulation tests, and the results show that the proposed method outperforms SVM and achieves comparable results.

## Data Availability

The data used to support the findings of this study are included within the article.

## Conflicts of Interest

The author does not have any possible conflicts of interest.

## Acknowledgments

The study was supported by the Social Science Foundation of Hunan Province (Grant No. 20WLH21).



## References

- [1] A. Boime, "The teaching reforms of 1863 and the origins of modernism in France," *Art Quarterly*, vol. 1, no. 1, pp. 1–39, 1977.
- [2] A. Howes, "Teaching reforms and the impact of paid adult support on participation and learning in mainstream schools," *Support for Learning*, vol. 18, no. 4, pp. 147–153, 2003.
- [3] K. Patwardhan, "How practical are the "teaching reforms" without "curricular reforms"?", *Journal of Ayurveda and integrative medicine*, vol. 1, no. 3, pp. 174–176, 2010.
- [4] J. R. Cowan, R. L. Light, B. E. Mathews, and G. R. Tucker, "English teaching in China: a recent survey," *TESOL Quarterly*, vol. 13, no. 4, pp. 465–482, 1979.
- [5] W. Yi'An, "English language teaching in China: trends and challenges," *TESOL Quarterly*, vol. 35, no. 1, pp. 191–194, 2001.
- [6] G. Hu, "English language teaching in China: regional differences and contributing factors," *Journal of Multilingual and Multicultural Development*, vol. 24, no. 4, pp. 290–318, 2003.
- [7] S. Y. Hsu, "Building language-learning environments to help technological university students develop English independent learning," *The Jalt Call Journal*, vol. 1, no. 2, pp. 51–66, 2005.
- [8] M. N. Islam, "Independent English learning through the Internet," *Journal of Language Teaching & Research*, vol. 2, no. 5, 2011.
- [9] N. Pekel, *Students' Attitudes towards Web-Based Independent Learning at Bilkent University School of English Language [Doctoral dissertation]*, Bilkent University, 2002.
- [10] M. T. Munir, S. Baroutian, B. R. Young, and S. Carter, "Flipped classroom with cooperative learning as a cornerstone," *Education for Chemical Engineers*, vol. 23, pp. 25–33, 2018.
- [11] F. Lestari, B. Saryantono, M. Syazali et al., "Cooperative learning application with the method of" network tree concept map": based on Japanese learning system approach," *Journal for the Education of Gifted Young Scientists*, vol. 7, no. 1, pp. 15–32, 2019.
- [12] B. Korbar, D. Tran, and L. Torresani, "Cooperative learning of audio and video models from self-supervised synchronization," 2018, <https://arxiv.org/abs/1807.00230/>.
- [13] J. S. Chen Hsieh, W. C. V. Wu, and M. W. Marek, "Using the flipped classroom to enhance EFL learning," *Computer Assisted Language Learning*, vol. 30, no. 1-2, pp. 1–21, 2017.
- [14] B. Ayçiçek and T. Yanpar Yelken, "The effect of flipped classroom model on students' classroom engagement in teaching English," *International Journal of Instruction*, vol. 11, no. 2, pp. 385–398, 2018.
- [15] Z. Turan and B. Akdag-Cimen, "Flipped classroom in English language teaching: a systematic review," *Computer Assisted Language Learning*, vol. 33, no. 5-6, pp. 590–606, 2020.
- [16] M. Y. Abdullah, S. Hussin, and K. Ismail, "Implementation of flipped classroom model and its effectiveness on English speaking performance," *International Journal of Emerging Technologies in Learning*, vol. 14, no. 9, 2019.
- [17] B. M. Smart and D. P. Saxon, "Online versus traditional classroom instruction: an examination of developmental English courses at an Alabama community college," *Community College Journal of Research and Practice*, vol. 40, no. 5, pp. 394–400, 2016.
- [18] Z. Wang, P. Zhang, W. Sun, and D. Li, "Application of data dimension reduction method in high-dimensional data based on single-cell 3D genomic contact data," *ASP Transactions on Computers*, vol. 1, no. 2, pp. 1–6, 2021.
- [19] Z. Chu, M. Hu, and X. Chen, "Robotic grasp detection using a novel two-stage approach," *ASP Transactions on Internet of Things*, vol. 1, no. 1, pp. 19–29, 2021.
- [20] J. Zhang, Y. Liu, H. Liu, and J. Wang, "Learning local-global multiple correlation filters for robust visual tracking with Kalman filter redetection," *Sensors*, vol. 21, no. 4, p. 1129, 2021.
- [21] W. Chu, P. S. Ho, and W. Li, "An adaptive machine learning method based on finite element analysis for ultra low-k chip package design," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, p. 1, 2021, In press.
- [22] P. Zheng, X. Wang, and J. Li, "Exploration and practice of curriculum ideological and Political Construction Reform—"take" information security" course as an example," *ASP Transactions on Computers*, vol. 1, no. 1, pp. 1–5, 2021.
- [23] L. Liang, Q. Yin, and C. Shi, "Exploring proper names online and its application in English teaching in university," *ASP Transactions on Computers*, vol. 1, no. 1, pp. 24–29, 2021.
- [24] Y. Tong, L. Yu, S. Li, J. Liu, H. Qin, and W. Li, "Polynomial fitting algorithm based on neural network," *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 1, pp. 32–39, 2021.
- [25] J. Zhang, J. Sun, J. Wang, and X. G. Yue, "Visual object tracking based on residual network and cascaded correlation filters," *Journal of ambient intelligence and humanized computing*, vol. 12, no. 8, pp. 8427–8440, 2021.
- [26] Y. Zhang, W. Li, L. Zhang, X. Ning, L. Sun, and Y. Lu, "AGCNN: adaptive gabor convolutional neural networks with receptive fields for vein biometric recognition," *Concurrency and Computation: Practice and Experience*, 2020, In press.
- [27] J. Zhang, X. Jin, J. Sun, J. Wang, and K. Li, "Dual model learning combined with multiple feature selection for accurate visual tracking," *IEEE Access*, vol. 7, pp. 43956–43969, 2019.
- [28] C. Yan, G. Pang, X. Bai et al., "Beyond triplet loss: person re-identification with fine-grained difference-aware pairwise loss," *IEEE Transactions on Multimedia*, p. 1, 2021, In press.

## Retraction

# Retracted: Deep Convolutional Neural Network and Weighted Bayesian Model for Evaluation of College Foreign Language Multimedia Teaching

### Wireless Communications and Mobile Computing

Received 1 August 2023; Accepted 1 August 2023; Published 2 August 2023

Copyright © 2023 Wireless Communications and Mobile Computing. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their

agreement or disagreement to this retraction. We have kept a record of any response received.

### References

- [1] T. Liu and L. Ning, "Deep Convolutional Neural Network and Weighted Bayesian Model for Evaluation of College Foreign Language Multimedia Teaching," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 1859065, 7 pages, 2021.

## Research Article

# Deep Convolutional Neural Network and Weighted Bayesian Model for Evaluation of College Foreign Language Multimedia Teaching

Tingting Liu<sup>1</sup> and Le Ning<sup>2</sup> 

<sup>1</sup>School of Foreign Languages, Ludong University, Yantai 264025, China

<sup>2</sup>School of Foreign Languages, Shandong Technology and Business University, Yantai 264005, China

Correspondence should be addressed to Le Ning; [ninglejilin@163.com](mailto:ninglejilin@163.com)

Received 16 June 2021; Revised 8 July 2021; Accepted 10 August 2021; Published 26 August 2021

Academic Editor: Shan Zhong

Copyright © 2021 Tingting Liu and Le Ning. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In colleges and universities, teaching quality evaluation is an integral part of the teaching management process. Many factors influence it, and the relationship between its evaluation index and instructional quality is complicated, abstract, and nonlinear. However, existing evaluation methods and models have flaws such as excessive subjectivity and randomness, difficulty determining the weight of indicators, easy over-fitting, slow convergence speed, and limited computing power, to name a few. Furthermore, the evaluation index system focuses primarily on teaching attitude, material, and methods, rarely taking into account preparation prior to teaching or the teaching situation throughout the teaching process, resulting in an incomplete evaluation. As a result, learning how to construct a model for objectively, truly, thoroughly, and accurately assessing the teaching quality of colleges and universities is beneficial not only to improving teaching quality but also to promoting scientific decision-making in education. This paper develops a teaching assessment model using a deep convolutional neural network and the weighted Naive Bayes algorithm. Based on the degree of influence of different characteristics on the assessment outcomes, a method to estimate the weight of each evaluation characteristic by employing the related probability of class attributes is proposed, and the corresponding weight is assigned for each evaluation index, resulting in a classification model ideal for teaching assessment that promotes standardization and intelligibility.

## 1. Introduction

With the continued development of higher education [1–4], determining how to fairly evaluate the teaching quality [5–7] of colleges and universities, promote the perfection of teaching objectives, and improve the teaching quality of colleges and universities is the key to furthering educational reform [8, 9], and it is also an urgent problem that needs to be solved now. As a result, evaluating teaching quality in colleges and universities has become an important part of the teaching management process [10], and researching methods or models for assessing teaching quality in colleges and universities has become a hot topic for scientific and standardized education and teaching.

Artificial neural network [11–14] is a nonlinear system [15] composed of many computational neurons which can be adjusted in different ways from layer to layer. It has the advantages of nonlinear ability, self-organization and self-learning ability, large-scale parallel processing, and so on. In the 1940s, McCulloch and Pitts [16] first proposed the mathematical model of neurons and became the forerunner in the study of artificial neural networks. Many new theories and algorithms of artificial neural networks have been proposed successively as a result of a large number of scholars joining the research, such as the perceptron model [17], back-propagation algorithm [18], Boltzmann machine [19, 20], unsupervised learning [21], and supervised learning [22–24], and their theoretical research and information

processing ability have improved and improved. Artificial neural networks have been applied to problems that cannot be solved by traditional methods and models as a mathematical model to deal with computation and have achieved good results in practice.

The back-propagation neural network (BP neural network) is a multilayer feedforward network that uses the back-propagation approach for training [25]. The goal of a BP neural network is to use the gradient descent method [26] to rectify error signals generated during forward propagation until the accuracy target is satisfied or the number of iterations is reached. Hecht-Nielsen [27] then showed that a BP network with a hidden layer may approximate a continuous function within any closed interval, meaning that any three-layer BP neural network can complete any mapping from  $N$  to  $M$  dimensions. Although the BP neural network has strong information processing and nonlinear mapping capabilities, it has some flaws, such as slow convergence speed and a tendency to fall into minimum values during the training process, making it ineffective in dealing with complex high-order abstraction problems. However, the multilayer perceptron with multiple hidden layers is a deep learning structure, which can be used to solve more complex and abstract nonlinear problems. After the publication of the research results of Geoffrey Hinton and Salakhutdinov [28], the academia and industry have set off a boom in the research of deep neural network, and the layer number and scale of neural network models have been greatly improved compared with the previous ones.

The main contributions of this paper are as follows:

- (1) To solve the problem of evaluating teaching quality in colleges and universities, promote continuous improvement of teaching goals, and promote scientific decision-making in education, this paper proposes a method based on a deep convolutional neural network and a weighted Bayesian model, all of which can help to improve teaching quality
- (2) This paper proposes to use convolutional neural networks to identify classroom behaviors and to use the weighted Bayesian incremental learning method to solve the problem of new samples arriving in batches. Through this strategy, it is not necessary to retrain the old sample data, and only need to adjust the model parameters according to the new sample data

The structure of this article is organized as follows. Section 2 presents the related work. Section 3 provides the details of the proposed method. Section 4 discusses the simulations and experiments, and Section 5 presents the conclusions.

## 2. Related Work

The United Kingdom, the United States, and Japan were among the first to begin and propose important ideas and methodologies, such as multiple intelligence theory, con-

structivism theory, and the Taylor evaluation model, in terms of teaching quality evaluation research. An American educationalist's book, *Introduction to Psychological and Social Measuring*, provided a theoretical foundation for the standardization of educational measurement and marked the maturation of educational measurement. In Russia, the most common assessment approaches are state-based assessment, school-based self-assessment, and self-supervision evaluation based on social competitiveness. The Japanese government proposes the "dual-track assessment model," which calls for a pluralistic, objective, and transparent evaluation system that incorporates both internal and external reviews. Currently, the teaching quality evaluation index system at colleges and universities is mostly focused on teaching attitude, material, and procedures, which are quantified as input feature vectors of applicable methodologies and models. Analytic hierarchy process, multivariate statistical analysis, fuzzy comprehensive evaluation, fuzzy hierarchy analysis, correlation analysis, ID3 algorithm, support vector machine, BP neural network model, and so on are some of the extant assessment methods and models. Literature [29] proposed an enhanced Apriori method for mining teaching system data to assess college and university teaching levels. A BP neural network is used to create a training quality evaluation model in the literature [30], demonstrating that the BP neural network approach is highly operational. It not only simplifies the evaluation process but also addresses some of AHP's flaws, such as subjectivity and randomness. Ge Chun et al. [31] used a genetic algorithm to select the best individual and set the optimal weight and threshold value for the BP neural network. Data training improved the accuracy of the classroom teaching quality rating.

## 3. Methodology

*3.1. Classroom Behavior Recognition Based on Convolutional Neural Network.* This section creates a classroom behavior recognition model using convolutional neural networks. The initial layer of the network model is the input layer of  $512 \times 512$  student classroom behavior photographs. The second, fourth, and sixth layers all use convolution, followed by the maximum pooling layer. Image features are extracted using the second to seventh layers. The full connection layer is the eighth and ninth layers, and the output layer is the last. 0.001 is the learning rate, RELU is the activation function, and BATCH\_SIZE is 100. The detailed network model structure is shown in Figure 1.

In the convolutional neural network model layout, the picture input is followed by the convolutional layer and the pooling layer, which alternate. After the convolution calculation, the convolutional layer will add the obtained result to the offset item and then activate the function with ReLU to get the feature map. Then, the dimension of the feature map will be reduced through maximum pooling. The convolution kernel is  $3 \times 3$ , and the maximum pooling runs through these convolution and pooling procedures three times before being input into the two full connection layers. In the full connection layer, nonlinear changes are also carried out at first, and then, the Dropout layer is

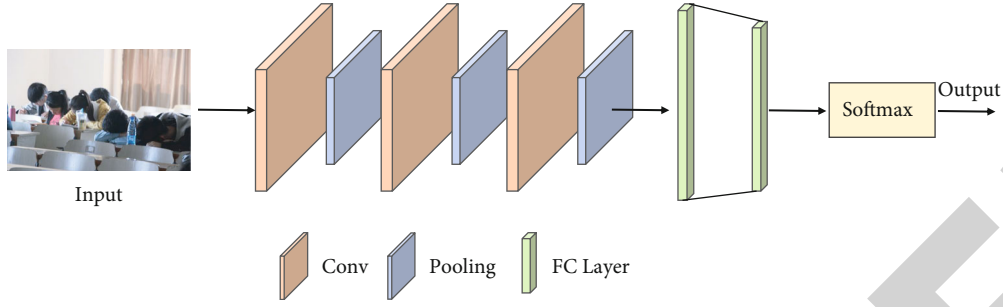


FIGURE 1: Schematic diagram of classroom behavior recognition model based on convolutional neural network.

added, and the loss rate is set at 0.4 to convert its feature graph into one-dimensional data. Finally, Softmax classifier is used to output the probability value of the predicted image.

### 3.2. Evaluation Classification Model Based on Weighted Naive Bayes

**3.2.1. Naive Bayes.** A classification algorithm based on Bayes' theorem is known as Bayesian classification. The basic premise of classification is to learn a significant amount of training data in order to estimate the prior probability of each category and, then, calculate the posterior probability of a certain instance  $X$  belonging to different categories, and finally determine the instance as the class with the largest posterior probability. Suppose  $D$  is the training dataset,  $A = \{A_1, A_2, \dots, A_n\}$  is the attribute variable set, and  $n$  is the number of attributes.  $C = \{C_1, C_2, \dots, C_m\}$  is the class variable set,  $m$  is the number of categories, then a training sample can be expressed as  $\{x_1, x_2, \dots, x_n, C_j\}$ ,  $C_j$  means that the class label of the training sample is known, and a test sample  $X$  can be expressed as  $\{x_1, x_2, \dots, x_n\}$ , judge the test sample. The probability of belonging to a certain category is calculated as follows:

$$p(C_j | X) = \arg \max_{C_j} \frac{p(X | C_j)p(C_j)}{p(X)}. \quad (1)$$

Naive Bayes (shown in Figure 2) is an effective classification algorithm in Bayesian classification methods. The classification model has the benefits of being simple to interpret, having a high computational efficiency, and being stable. In some cases, it outperforms decision-making and Classifiers like tree and SVM.

The root node  $C$  is the class variable, and the leaf node  $\{A_1, A_2, \dots, A_n\}$  is the attribute variable. The generic Bayes classification model is based on the Naive Bayes classification model, which does not impose attribute independence restrictions. In practice,  $p(X)$  is usually a constant, so the calculation equation of Naive Bayes is as follows:

$$p(C_j | X) \propto \arg \max_{C_j} p(X | C_j)p(C_j), \quad (2)$$

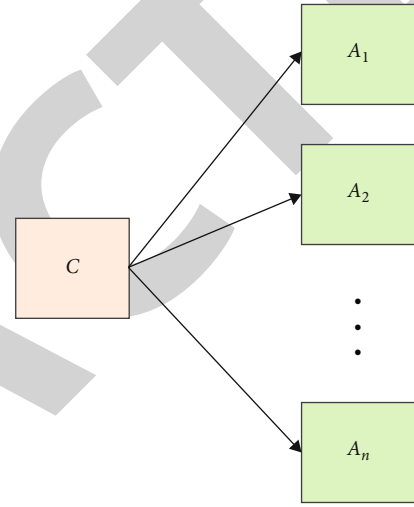


FIGURE 2: Schematic diagram of Naive Bayes.

where  $p(C_j)$  is the class prior probability, which can be learned through training data. The calculation equation is

$$p(C_j) = \frac{S_j}{S}, \quad (3)$$

where  $S_j$  represents the number of class  $C_j$  in the training sample and  $S$  represents the total number of training samples.

According to the assumption of conditional independence, the calculation formula of  $p(X | C_j)$  can be simplified as

$$p(X | C_j) = \prod_{i=1}^n p(x_i | C_j). \quad (4)$$

**3.2.2. Evaluation Attribute Weights Based on Weighted Naive Bayes.** In this paper, the Weighted Naive Bayes (WNB) classification algorithm is used to assign a reasonable weight to attributes based on their contribution to classification, which not only keeps the Naive Bayes algorithm fast but also reduces the impact of the attribute conditional independence assumption on the classifier's performance. The following is the formula for calculating it:

$$p(C_j | X) = \arg \max_{C_j} p(C_j) \prod_{i=1}^n p(A_i | C_j)^{w_i}, \quad (5)$$

TABLE 1: Comparative experiment results on the GTZAN dataset.

Times	1	2	3	4	5	6	7	8	9	10
SVM	0.71	0.70	0.71	0.69	0.71	0.71	0.73	0.68	0.71	0.70
NB	0.68	0.71	0.70	0.72	0.70	0.72	0.70	0.71	0.72	0.71
Ours (WNB)	0.75	0.76	0.73	0.74	0.73	0.74	0.74	0.73	0.74	0.75

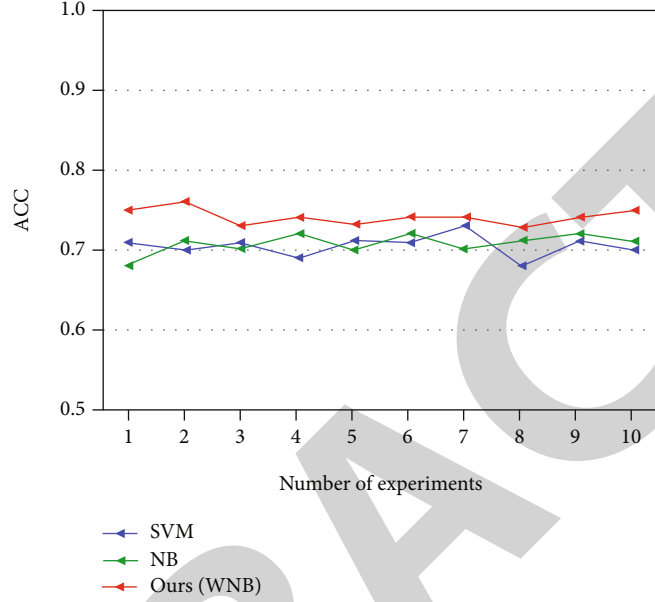


FIGURE 3: Comparison result of evaluation accuracy.

where  $w_i$  represents the weight of attribute  $A_i$ , which determines the importance of different attributes in the classification process. The larger the value of  $w_i$ , the more important the corresponding attribute  $A_i$  is for classification.

Assuming a specific instance  $X$ , when the attribute  $A_i$  of  $X$  takes the value  $a_k$ , for the category  $C_j$ , the calculation formula of the correlation probability  $p(A_i | rel)$  and the irrelevant probability  $p(A_i | norel)$  of the attribute  $A_i$  with respect to  $C_j$  is as follows:

$$p(A_i | rel) = \frac{\text{count}(A_i = a_k \wedge C_j)}{\text{count}(A_i = a_k)}, \quad (6)$$

$$p(A_i | norel) = 1 - p(A_i | rel), \quad (7)$$

where count represents the statistical number. When the attribute  $A_i$  value is  $a_k$  and belongs to the  $C_j$  category, the attribute weight calculation formula is as follows:

$$w(A_i, a_k, j) = \frac{p(A_i | rel)}{p(A_i | norel)}. \quad (8)$$

Therefore, the specific calculation formula of the WNB classification algorithm is as follows:

$$p(C_j | X) = \underset{C_j}{\operatorname{argmax}} p(C_j) \prod_{i=1}^n p(A_i | C_j)^{w(A_i, a_k, j)}. \quad (9)$$

Finally, based on the specific value of each characteristic, the weight value of the likelihood associated with the current category label is chosen for computation, and the result value of each category is compared. The greatest value corresponds to the highest category in the classification.

## 4. Experiments and Results

**4.1. Experimental Setup.** The hardware configuration of the experimental environment is as follows: CPU is Intel E5-1607V3 quad-core 3.1GHz, graphics card is RTX2070 (8G), memory is 16G, and operating system is Windows10. The experimental simulation uses a framework based on TensorFlow 1.9.0, and the programming language is implemented using Python 3.5 and MATLAB R2017b. The learning rate of the model in this paper is 0.01, the number of iterations is 2000, and the batch size is 10.

**4.2. Dataset.** The student evaluation data used in the experiment comes from real data in the educational administration management system of a university. The back-end database management system of the educational administration management system of the school uses Oracle as the database management system. The student evaluation database of the school stores the data from the second half of 2004. Up to the first half of 2020, there are a total of 16 school years and 32 semesters of all teaching evaluation data, and a total of 222,190 student evaluation records. Student

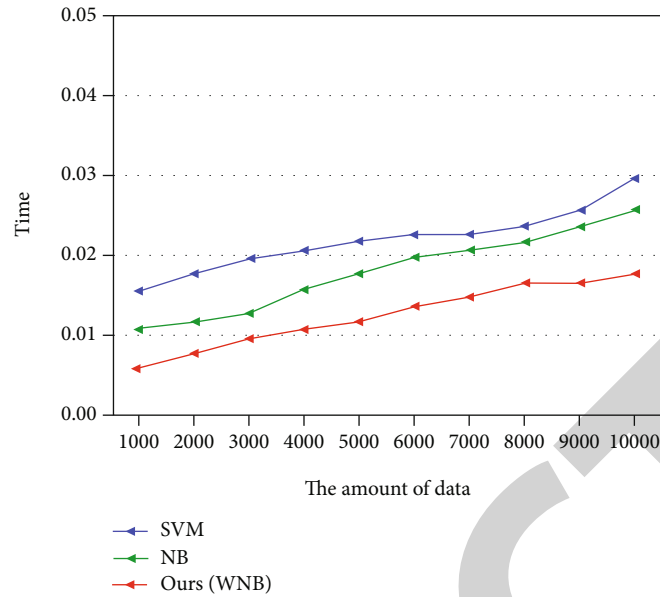


FIGURE 4: Comparison result of time.

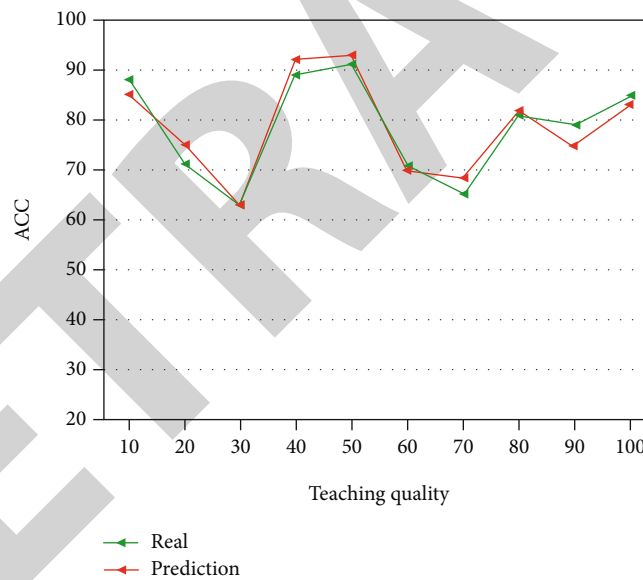


FIGURE 5: Comparison result of the prediction and real.

evaluation of teaching is organized and implemented in the form of an online evaluation of teaching. In order to ensure that every student must participate in the evaluation of teaching, the system has adopted a mandatory treatment. The system requires that the evaluation of teaching by students must be conducted in the time after the end of the course and before the examination of the course; otherwise, the scores of students who did not participate in the evaluation of the course refuse to enter the system.

**4.3. Evaluation Index.** This chapter determines the classification accuracy rate (Acc). The following is the procedure for calculating classification accuracy:

$$\text{Acc} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \quad (10)$$

**4.4. Experimental Results.** To conduct cross-validation testing, we randomly selected 70% of the data as the training set and 30% of the data as the test set. The classification accuracy of the NB and WNB algorithms was determined using ten cross-validation trials. The individual experimental outcomes are shown in Table 1 below.

As demonstrated in Figure 3, the average classification accuracy of the Naive Bayes algorithm on this dataset is 0.707, while the weighted naive Bayes approach is 0.741. The weighted naive Bayes algorithm has a greater classification

TABLE 2: Comparative experiment results on the GTZAN dataset.

Times	1	2	3	4	5	6	7	8	9	10
$lr = 0.1$	0.74	0.75	0.71	0.73	0.71	0.73	0.72	0.72	0.73	0.72
$lr = 0.001$	0.63	0.75	0.72	0.73	0.71	0.71	0.69	0.71	0.73	0.70
Ours ( $lr = 0.01$ )	<b>0.75</b>	<b>0.76</b>	<b>0.73</b>	<b>0.74</b>	<b>0.73</b>	<b>0.74</b>	<b>0.74</b>	<b>0.73</b>	<b>0.74</b>	<b>0.75</b>

accuracy than the regular naive Bayes algorithm on the instructional evaluation dataset.

It can be seen from Figure 4 that because the incremental model does not need to retrain and calculate the previously trained dataset, it only needs to classify and calculate the increased data, directly merge with the previous training value, and update the relevant parameters of the model, saving It saves time and improves the efficiency of the classification model. In addition, Figure 5 depicts a comparison between the predicted and actual results. The model's usefulness is demonstrated by the experimental results.

It can be seen from Table 2 that by setting the ablation experiment, we can clearly see that when the learning rate is set to 0.01, the model achieves the best performance. Therefore, we can determine the effectiveness of the hyper-parameters of the model in this paper.

## 5. Conclusion

In colleges and universities, teaching quality evaluation is an integral part of the teaching management process. It is not only beneficial to increase teaching quality but also to encourage scientific decision-making in education, to learn how to develop an objective, actual, comprehensive, and accurate model to evaluate teaching quality in colleges and universities. This research develops a teaching assessment model using a deep convolutional neural network and the weighted Naive Bayes method. A method to estimate the weight of each evaluation characteristic by employing the related probability of class attributes is proposed based on the degree of influence of different characteristics on the assessment outcomes, and the corresponding weight is set for each evaluation index, so as to construct a classification model suitable for teaching evaluation, which is conducive to promoting the standardization and intelligence of teaching management in colleges and universities. We conducted a series of simulation, comparison, and ablation experiments. The results of the comparison experiments show that the method in this paper has achieved competitive performance. In addition, through ablation experiments, this paper further confirms the validity and superiority of the model in this paper.

## Data Availability

The data used to support the findings of this study are included within the article.

## Conflicts of Interest

All the authors do not have any possible conflicts of interest.

## Acknowledgments

This study was supported by the Shandong Social Science and Research Foundation of China (Grant No.17CWZJ07\18CWZJ34).

## References

- [1] G. Dawe, R. Jucker, and S. Martin, "Sustainable development in higher education: current practice and future developments. A report to the Higher Education Academy, York (UK)," 2005, <https://www.heacademy.ac.uk/system/files/sustdevinHEfinalreport.pdf>.
- [2] M. Svanström, F. J. Lozano-García, and D. Rowe, "Learning outcomes for sustainable development in higher education," *International Journal of Sustainability in Higher Education*, vol. 9, no. 3, pp. 339–351, 2008.
- [3] M. Barth, J. Godemann, M. Rieckmann, U. Stoltenberg, M. Rieckmann, and U. Stoltenberg, "Developing key competencies for sustainable development in higher education," *International Journal of Sustainability in Higher Education*, vol. 8, no. 4, pp. 416–430, 2007.
- [4] E. Abad-Segura and M. D. González-Zamar, "Sustainable economic development in higher education institutions: a global analysis within the SDGs framework," *Journal of Cleaner Production*, vol. 294, article 126133, 2021.
- [5] A. Arrona-Palacios, K. Okoye, C. Camacho-Zuñiga et al., "Does professors' gender impact how students evaluate their teaching and the recommendations for the best professor?," *Heliyon*, vol. 6, no. 10, article e05313, 2020.
- [6] D. M. Dockterman, *Discrepancies between Students' and Teachers' Ratings of Instructional Practice: A Way to Measure Classroom Intuneness and Evaluate Teaching Quality*, [Ph.D. thesis], UCLA, 2017.
- [7] D. Feistauer and T. Richter, "How reliable are students evaluations of teaching quality? A variance components approach," *Assessment & Evaluation in Higher Education*, vol. 42, no. 8, pp. 1263–1279, 2017.
- [8] R. Nasser, "Qatar's educational reform past and future: challenges in teacher development," *Open Review of Educational Research*, vol. 4, no. 1, pp. 1–19, 2017.
- [9] V. Sahuichenko, V. Shynkarenko, I. Bezena, O. Perederii, and O. Martynenko, "State policy on the formation of students' civic and social competences in conditions of educational reform," *Revista Educação & Formação*, vol. 5, no. 3, article e3080, 2020.
- [10] S. Samuel, "A conceptual framework for teaching management accounting," *Journal of Accounting Education*, vol. 44, pp. 25–34, 2018.
- [11] X. Ning, Y. Wang, W. Tian, L. Liu, and W. Cai, "A biomimetic covering learning method based on principle of homology continuity," *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 1, pp. 9–16, 2021.



## Research Article

# A Mental Health Assessment Model of College Students Using Intelligent Technology

Keke Li  and Weifang Yu 

Wuxi Vocational College of Science and Technology, Wuxi 214028, China

Correspondence should be addressed to Weifang Yu; 1301806@wxsc.edu.cn

Received 14 July 2021; Revised 7 August 2021; Accepted 10 August 2021; Published 26 August 2021

Academic Editor: Shan Zhong

Copyright © 2021 Keke Li and Weifang Yu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

College students are under increasing competition pressure, which has a negative impact on their mental health, as the pace of learning and life accelerates, as well as the increasingly difficult employment situation. As a result, emphasizing the importance of college students' mental health and fully addressing it has become a top priority in the work of colleges and universities. However, some students and even teachers are currently unconcerned about mental illness, making it difficult for students with psychological abnormalities to receive timely detection and effective treatment. As a result, it is the responsibility of student management for colleges and universities to identify and intervene early in the mental health problems of college students. Through the use of multimodal data and neural network models, it is now possible to evaluate and predict the mental state of college students in real time, thanks to the advancement of intelligent technology. Therefore, a novel multimodal neural network model is proposed in this paper. Our model is divided into two branches in particular. The traditional mental health assessment and prediction algorithm, which is based on the improved BP neural network and the International Mental Health Scale SCL-90, is one of the branches. Given how difficult it is to meet the requirements for the accuracy of college students' mental health assessments using this method, our other branch is computer vision-based facial emotion recognition of college students, which is used to aid in the evaluation of mental health assessments. Our model demonstrates competitive performance through simulation and comparative experiments.

## 1. Introduction

With today's fierce competition and increasing pressures in life, college students' mental health problems [1–3] have become more visible, and their mental health conditions [4–6] are concerning. People with severe mental disorders or mental illnesses are forced to suspend school, drop out of school [7, 8], self-harm [9, 10], commit suicide, and even break the law in an endless stream among college students. It is critical and urgent to improve college students' overall quality, particularly their psychological quality [11], cultivate exceptional social talents, improve mental health education, and predict mental health.

College students are outstanding members of the youth population, representing a high intellectual group, and their mental health is critical. College students are in a critical tran-

sition period in their development and maturity. They will face a variety of issues during this time, including emotions and socialization [12, 13]. If they are not handled properly, they can lead to depression, anxiety, and other psychological issues. This is extremely harmful to college students' development. It is not uncommon to come across examples of exceptional college students who failed to deal with the final suicide due to emotional issues [14].

College students are confronted with a more complicated environment as society evolves. On the one hand, the steady increase in the number of graduates has put more pressure on college students to find work. Advances in science and technology, on the other hand, have increased the barriers to employment and education, and college students are facing greater challenges. Furthermore, college students face pressures in the areas of learning, communication,

emotions, and life. Such complex relationships are causing an increase in psychological problems [15]. Psychological problems affect all aspects of college students, not only have an impact on their studies and employment but may seriously endanger their own health. A survey conducted by the Chinese Center for Disease Control and Prevention showed that 16.0%-25.4% of college survivors have general mental health problems such as mild anxiety and stress, of which about 2.8% have mental health problems of varying degrees, and some college survivors have varying degrees of mental health. Accidents of grief occur in severe cases. According to a survey on student mental health, 80% of students said they had experienced psychological problems, 11% of students said they were in an unhealthy mental state, and 35% of students said they were under increased psychological pressure at school. According to the survey, psychological issues have surpassed financial difficulties as the leading cause of college dropout. The recurrence of campus tragedies demonstrates that mental health issues among college students have progressed to the point where they are now life-threatening. In this context, we should investigate the assessment and prediction of college students' mental health [16–18].

In recent years, with the rapid development of intelligent technologies [19–21] such as neural networks [22–24], it has extraordinary performance for nonlinear problems such as college students' mental health assessment. Therefore, the proposed model in this paper is mainly based on convolutional neural networks and improved BP neural networks [25]. The neural network and Bayesian method are combined to learn by constructing a suitable training model [26] from the selected typical samples, selecting the best training situation to grasp the internal relationship between the input and output, and obtaining the network weight vector of the knowledge about the problem. Apply the obtained prior probability to the Bayesian formula and combine the conditional probability obtained by actual statistics to calculate the impact of various indicators on the mental health of college students. Secondly, we also recognize the facial emotions of college students based on computer vision [26, 27] calculations to further improve the performance of college students' mental health assessment.

The main innovations of this article are as follows:

- (1) This paper proposes a novel dual-branch neural network model based on intelligent technology for college students' mental health assessment. One branch is based on improved BP neural network for mental health prediction, and the other branch is based on computer vision-based facial emotion recognition for college students
- (2) This paper proposes to combine neural networks with Bayesian methods, by constructing a suitable training model from selected typical samples, selecting the best training situation to grasp the internal relationship between input and output, and obtaining network rights about the knowledge of the problem. The vector is learned and then converted into

the weight of each indicator according to the formula, that is, the prior probability. Apply the obtained prior probability to the Bayesian formula and combine the conditional probability obtained by actual statistics to calculate the influence of various indicators on the mental health of college students

- (3) Small scale is the key to facial emotion recognition in complex environments. We first analyze the scenes of small scale faces, make reasonable structural adjustments to the classic classification network, and introduce attention mechanism modules to improve the subtleties of the network model. Feature extraction capabilities to improve the accuracy of facial emotion recognition

The rest of the paper is arranged as follows: the second section presents the related work, the third section proposes the methodology, the fourth section provides the experiments and results, and finally, the fifth section proposes the conclusions.

## 2. Related Work

*2.1. Mental Health-Related Research.* The American Jiehata advocates the “positive mental health” health concept as the most authoritative from a psychological standpoint. Jiehata believes that even in the face of adversity and setbacks, people can maintain a stable and stable mental state and maximize their performance. Maintain self-control and rationality, deal calmly with pressure and stimuli, and clarify your personal development path. Mental health, according to the famous psychologist Sigmund Freud, is defined as the ability to love and work. Mentally healthy people, he believes, should be sensible, loving, able to maintain close relationships with others, have a sense of self-worth, accept reality, and have peace of mind. Freud believes that people's mental health is related to their early experiences. The early years of college students are usually experienced in the family; so, family relations, family economic conditions, and other factors will have a good or bad impact on college students' mental health; Kara Zwin uses the Internet to track the life and study of American college students, records the behavioral characteristics of mentally unhealthy students according to different levels of mental health, and finds that the psychological problems of college students are complex and changeable. Many college students suffer from multiple psychological problems, and most students suffer from at least one. Holly Anne followed up with 97357 college students and found that the weight of college students is a factor that needs attention, because in the research, college students will cause psychological problems because of their weight, and the weight of students with psychological problems is more likely to be affected.

*2.2. Facial Emotion Recognition.* In recent years, facial emotion recognition [28–30] has also attracted a lot of attention and has become a hot research direction. Facial emotion recognition is divided into static emotion recognition and dynamic emotion recognition according to the feature

representation. Static only needs to consider the characteristics of the current image, and dynamic needs to consider the relationship between adjacent frames in the video. For the emotion recognition of static images, it is divided into two parts at the beginning. First, SIFT, LBP, and other operators are used to extract the features, and then the classifier is used to complete the classification, as shown in Figure 1. Fasel [31] et al. found that the ability of convolutional neural network to extract features is better than that of multilayer perceptrons when the position of the face changes and the scale changes greatly. Matsugu et al. [32] and others used a convolutional neural network model to solve the problem of face position and scale changes in facial emotion recognition. Yao et al. [33] et al. proposed the HoloNet network model to achieve end-to-end classification. In the shallow design, the phase convolution module was proposed, and the two-way activation method was introduced. Under the premise of ensuring the output dimension remains unchanged, the number of convolution kernels was changed. Halve, while retaining the positive and negative phase modulus information; in the middle-level design, the phase residual module is proposed, which combines the two-way activation method with the residual structure to increase the depth of the network and retain the positive and negative phase modulus information. In the deep design, the initial residual module is proposed, by designing a wider and deeper combination structure than the Inception structure [34] and introducing a multiscale deep feature extraction and fusion mechanism. Zhao et al. [35] et al. proposed a cascaded network structure; that is, the network is first trained to detect human faces, then the areas related to facial expressions are detected hierarchically, and finally, emotion recognition is performed based on these areas.

A recurrent neural network model is typically used to realize emotion recognition of dynamic sequences. To complete the two tasks of facial emotion recognition in static images and dynamic sequences, Sun et al. [36] proposed a method based on the fusion of traditional features and deep features: extract MSDF, DCNN, and RCNN features from static images and then use SVM classifier, to classify and then use the fusion network to arrive at a final classification result. Zhang et al. [37] proposed the MSCNN network for detecting facial feature points and sending them to the PHRNN recurrent neural network for facial expression classification.

### 3. Methodology

**3.1. Definition of Mental Health.** College students' mental health is extremely important to their physical well-being, and it is an important aspect of their overall health. There are corresponding standards for human physical health, and mental health has its own set of criteria for evaluation, but the main content of mental health standards differs significantly from physical health standards. Even if a person's physiology is perfect, his mental health may not be, and mental illnesses may arise. Only by grasping the concept of mental health can you understand your own mental health problems, so as to adopt targeted training and treatment to reach the level of mental health.

Mental health is a broad and complex concept that encompasses a variety of fields, including medical, psychological, and social phenomena. As a result, different scholars have different perspectives and opinions on how to manage health. Clement believes that mental health encompasses not only physical well-being but also the psychological level of happiness. In this regard, he proposed that mental health should be classified according to the degree of happiness experienced by the individual. This level should range from a low happiness state to a medium state and then to the continuum of the most energetic and higher state. Bellonio proposed that mental health includes three aspects: emotional, psychological, and social well-being. Emotional happiness refers to the degree of individual's response to positive and negative emotions, while psychological happiness refers to the establishment and development of potential interpersonal relationships and the degree of acceptance of individuals in the process of pursuing their own meaning and goals in life. Social happiness is that an individual identifies with the society and its parts and obtains satisfaction from his contribution to the society or part of the society.

The Third International Mental Health Conference held in 1946 defined mental health as follows: mental health refers to the development of an individual's mental state to the best state within the range of physical, intellectual, and emotional not contradicting the psychology of others.

#### 3.2. Improved BP Neural Network Based on Bayesian

**3.2.1. BP Network.** The basic processing unit of the BP neural network is neurons, which are made up of a large number of them. Each layer of neurons affects only the next layer, and the same layer does not interact with the others. The weights represent the degree of influence of the upper layer neurons on the lower layer neurons. The neural network's training process involves continuously adjusting the weight threshold between neurons based on the training sample set of system input and target output data in order to establish the appropriate mathematical relationship for the input and output data and express their inherent characteristics that meet the target requirements. The input layer and the output layer each have one layer, and the hidden layer can be set at one or more layers according to actual needs. The number of hidden layer nodes is determined by repeated comparisons based on experience and experimental results. The topological structure of a typical three-layer feedforward BP network is shown in Figure 2.

The forward propagation of the working signal and the backward propagation of the error signal are the two main parts of the BP network training process. The first step is performed if the variance between the actual output obtained by the output layer, and the target output does not meet the target accuracy. The error signal is transmitted back according to the previous distance in the second stage, and the network's weights and thresholds are continuously corrected. For the next iteration, the adjusted weights are used, and the error square between the actual output of the network and the target output is squared to achieve the required accuracy.

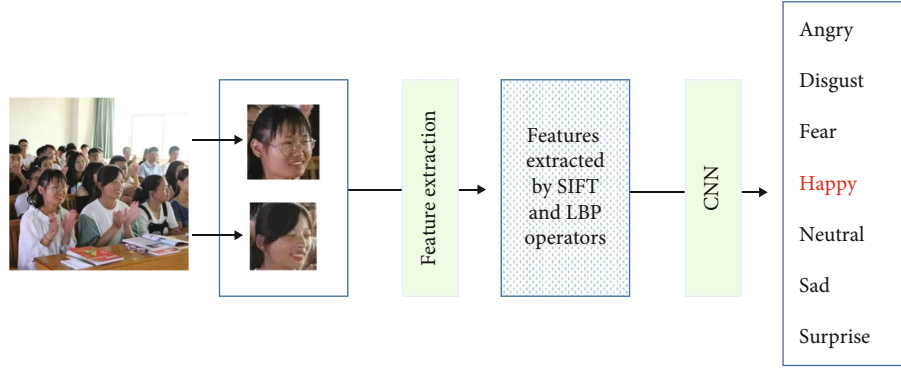


FIGURE 1: Schematic diagram of static emotion recognition. The matrix after feature extraction represents the features extracted by SIFT and LBP operators.

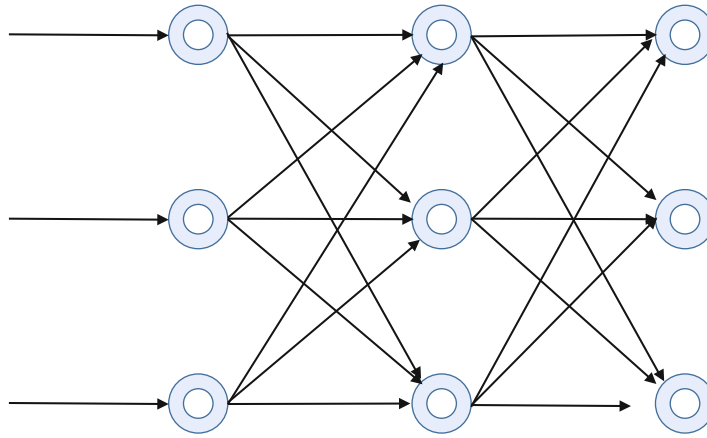


FIGURE 2: Schematic diagram of the BP neural network structure.

Assign a random value in the interval  $(-1, 1)$  to each connection weight  $w_{ij}$ ,  $v_{jt}$ , and threshold  $j$ ,  $tr$  of the network. For the input layer, the input and output are the same; then,

$$\theta_{X_k} = x_k = (X_1^k, X_2^k, \dots, X_n^k). \quad (1)$$

For the hidden layer, first obtain the input  $a_j$  of each neuron in the hidden layer according to the input sample data  $x_k = (X_1^k, X_2^k, \dots, X_n^k)$ ,  $w_{ij}$ , and  $\theta_j$  and then use  $a_j$  to find the output  $b_j$  of each neuron:

$$a_j = \sum_{i=1}^n w_{ij} x_i - \theta_j, \quad (2)$$

$$b_j = f(a_j). \quad (3)$$

For the output layer, calculate the output  $L_t$  of each neuron in the output layer through the connection weight  $v_{jt}$ , threshold  $r_t$ , and hidden layer output  $b_j$  of each neuron in the hidden layer and then calculate the actual output  $C_t$  of the output layer through the transfer function:

$$L_t = \sum_{j=1}^p v_{jt} b_j - r_t, \quad (4)$$

$$C_t = f(L_t). \quad (5)$$

Then, we calculate the error, and the error of the output layer and the hidden layer are as follows:

$$\delta_{pj} = (y_{pj} - O_{pj}) O_{pj} (1 - O_{pj}), \quad (6)$$

$$\delta_{pj} = O_{pj} (1 - O_{pj}) \sum_j \delta_{pj} W_{pj}. \quad (7)$$

Then, correct the error of the connection weight matrix between the upper and lower layers of the BP neural network:

$$W_{ij}(n+1) = \eta O_{pi} \delta_{pj} + W_{ij}(n), \quad (8)$$

where  $\eta$  represents the learning speed.

Due to the slow convergence speed of the BP network and the poor generalization ability, we need to make basic improvements to the BP network.

**3.2.2. Optimization of the BP Network.** Because the standard algorithm of the BP network that uses the steepest gradient descent method to adjust the weight has many defects such as too slow convergence and easy to fall into local minima, we use the LM algorithm for optimization.

The LM algorithm is an improved algorithm that integrates the local convergence of the quasi-Newton method and the global advantages of the gradient descent method to adaptively adjust the network weights and thresholds. Let  $w^k$  and  $w^{k+1}$  be divided into vectors composed of the weights and thresholds of iteration  $k$  and  $k+1$ ,  $w = (w_1, w_2, \dots, w_m)$ , where  $m$  is the number of weight thresholds  $w^{k+1} = w^k + \Delta w$ ; Then,

$$\Delta w = -[\nabla^2 E(w)]^{-1} \nabla E(w), \quad (9)$$

where  $\nabla E(w)$  and  $\nabla^2 E(w)$  are the first derivative and the second derivative of the error function, respectively, and  $E(w)$  is the error function.

$$E(w) = \frac{1}{2} e^{-T} e = \frac{1}{2} \sum_{j=1}^q \sum_{k=1}^p e_{jk}(w)^2, \quad (10)$$

where  $e$  is the error vector  $e = [e_{11}, e_{12}, \dots, e_{1p}, e_{21}, \dots, e_{q1}, e_{q2}, \dots, e_{qp}]$ ,  $q$  is the number of input sample groups,  $p$  is the number of output target vectors, and  $n = p * q$  is the dimension of the error vector. Then, the first derivative of the error function is

$$\nabla E(w) = \sum_{i=1}^n \frac{\partial E(w)}{\partial w_i} = \sum_{i=1}^n \sum_{j=1}^q \sum_{k=1}^p \frac{\partial e_{jk}(w)}{\partial w_i} e_{jk}(w). \quad (11)$$

And the second derivative is

$$\nabla^2 E(w) = \sum_{i=1}^n \sum_{j=1}^q \sum_{k=1}^p \sum_{i=1}^n \left( \frac{\partial e_{jk}(w)}{\partial w_i} \frac{\partial e_{jk}(w)}{\partial w_i} + e_{jk}(w) \frac{\partial^2 e_{jk}(w)}{\partial w_i \partial w_i} \right). \quad (12)$$

Then, we can get

$$\nabla E(w) = D^T(w) e(w), \quad (13)$$

$$\nabla^2 E(w) = D^T(w) D(w) + S(w), \quad (14)$$

where  $S(w)$  and  $D(w)$  are the Jacobian matrix, and the calculation equation of the matrix is as follows:

$$D(w) = \begin{bmatrix} \frac{\partial e_{11}(w)}{\partial w_1} & \frac{\partial e_{11}(w)}{\partial w_2} & \dots & \frac{\partial e_{11}(w)}{\partial w_n} \\ \frac{\partial e_{12}(w)}{\partial w_1} & \frac{\partial e_{12}(w)}{\partial w_2} & \dots & \frac{\partial e_{12}(w)}{\partial w_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial e_{qp}(w)}{\partial w_1} & \frac{\partial e_{qp}(w)}{\partial w_2} & \dots & \frac{\partial e_{qp}(w)}{\partial w_n} \end{bmatrix}. \quad (15)$$

Let  $H = D^T D$ , and then  $H$  is the Hessian matrix. Then, the calculation equation of the LM algorithm is as follows:

$$\Delta w = -[D^T(w) D(w) + \mu I]^{-1} D(w) e(w). \quad (16)$$

**3.2.3. Bayesian-BP Branch Network.** We combine the Bayesian and BP network to analyze and judge the multiple factors that cause college students' mental health diseases and provide effective guidance and reference for the prevention of college students' mental health diseases.

First, perform feature screening and feature engineering on the collected college student mental health big data. The training sample converges quickly in the learning process and reaches the required accuracy of the target and the prediction effect of the test sample. There is a small difference from the actual expected target, that is, when the accuracy rate is high. Transform the weight vector output by the network to obtain the prior probability and apply it to the Bayesian method to obtain the posterior probability. Sort the obtained posterior probability, that is, the degree of influence of each physiological index on the mental illness of college students, combined with computer vision-assisted facial emotion recognition, etc., finally, get the most important factors among them. The schematic diagram of the branch network is shown in Figure 3.

**3.3. Vision-Based Branch Network.** Due to the lack of accuracy caused by subjective uncertain factors in the prediction based on the self-rating symptom scale SCL90, this paper innovatively considers the intelligent method based on computer vision. But in complex environments such as classrooms, the emotion recognition of small-scale faces is the key to the branch network in this section. In this article, we analyze the scenes of small-scale faces and make reasonable structural adjustments to the classic classification network and introduce the attention mechanism module to improve the subtle feature extraction ability of the network model, thereby improving the accuracy of facial emotion recognition.

**3.3.1. VGG Model.** We choose VGG as the backbone network, as shown in Figure 4. For ordinary classification tasks, VGG has a strong feature extraction ability and can usually achieve good results. For fine-grained classification tasks, the model is required to have very high requirements for the ability to extract subtle features, and its performance is average. Face emotion recognition is to classify a variety of emotions. There are many expressions that seem to have very small differences, such as disgust and fear, sadness,

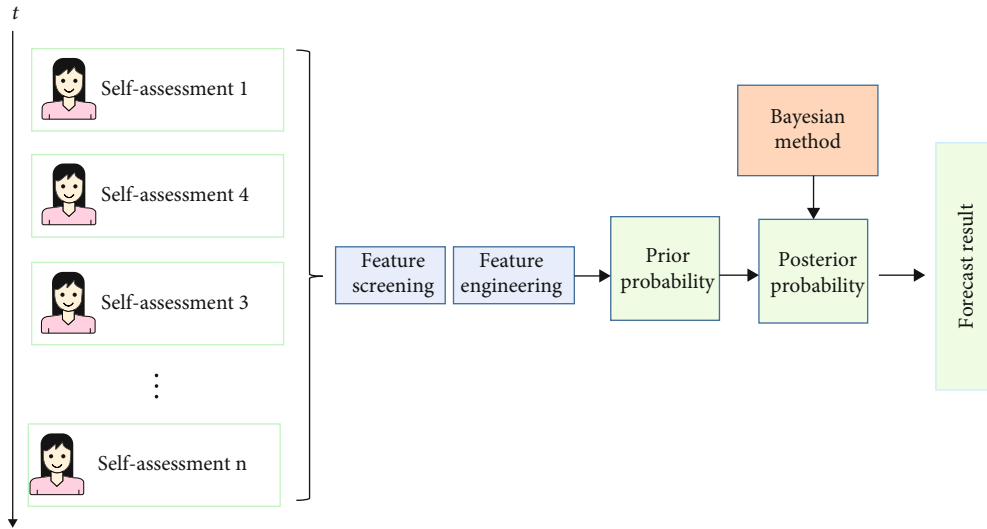


FIGURE 3: Schematic diagram of the Bayesian-BP branch network.

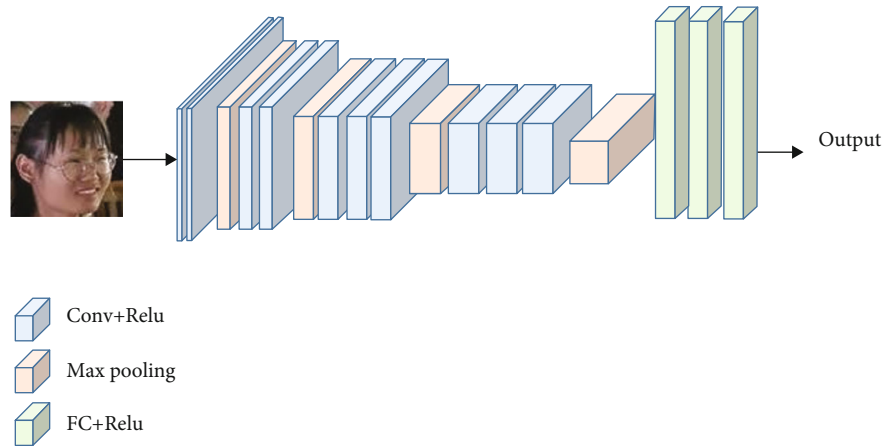


FIGURE 4: Schematic diagram of the VGG network.

and neutrality, which can be counted as fine-grained classification tasks. Combining the characteristics of the scene and the model, the network structure should be adjusted to adapt to the  $50 \times 50$  low-resolution input used, and the model needs to be compressed to alleviate the overfitting situation.

**3.3.2. Attention Module.** The discrimination of facial expressions is very small in facial emotion classification tasks, and it can be divided into fine-grained classification. The most common supervised method of fine-grained classification divides the image into several areas, which are then fed into a composite structure network along with the original image. This can often produce good results for high-resolution images, but there are issues such as which area to choose and how large the area should be; however, it is difficult to achieve for low-resolution images, primarily because the target area cannot be obtained.

Inspired by the above focus on a certain area, this paper proposes to use the attention mechanism [38] module (Figure 5) to achieve the above goals. The attention mecha-

nism module allows the model to learn and focus on important information areas autonomously during the training process, eliminating the need to divide areas and other steps. The attention mechanism module includes two parts: channel attention module and spatial attention module. Given a three-dimensional feature map [39], compress it into a two-dimensional feature map according to the channel and spatial dimensions, then connect it with convolution or full connection, and finally restore it to a weight between 0 and 1 for each pixel point.

**3.3.3. Vision-Based Model.** We embed the attention mechanism module into the VGG network to construct a branch network of emotion recognition for college students based on vision, as shown in Figure 6.

**3.3.4. Our Model.** Finally, we merged the Bayesian-BP network branch and the vision-based branch to construct a mental health assessment model for college students, as shown in Figure 7.

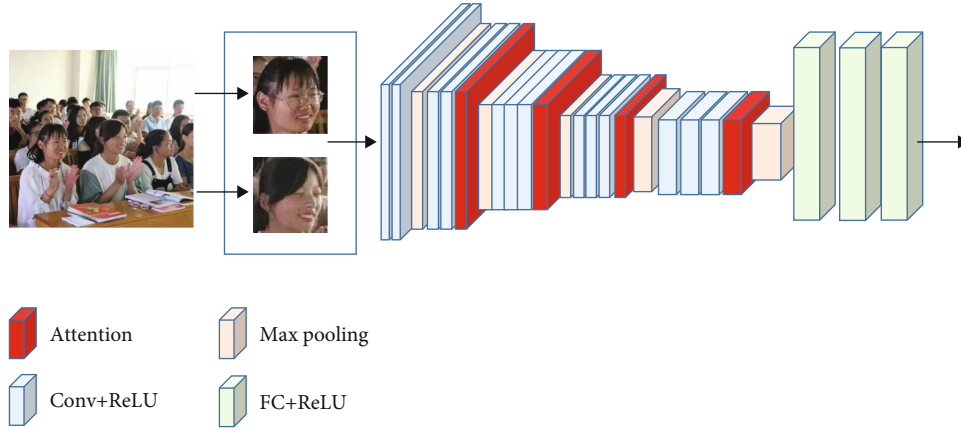


FIGURE 5: Schematic diagram of the VGG network.

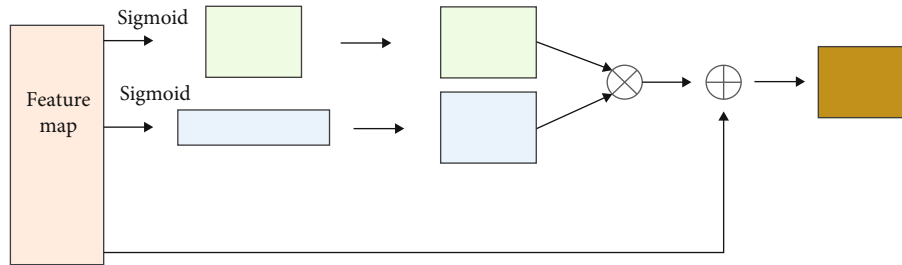


FIGURE 6: Schematic diagram of the VGG network.

## 4. Experiments

**4.1. Experimental Environments.** This article uses the deep learning framework PyTorch developed by Facebook to build and train a convolutional neural network model. The environment configuration is shown in Table 1. We use 80% of the samples of the data set as the training set and 20% of the samples as the test set. In addition, the batch size is 100, and the number of iterations is 5000.

**4.2. Dataset.** In this paper, more than 2000 cases of college students' mental health data are collected from a university, and the unreasonable parts are filtered out. 1877 cases of valid data are selected as the training sample set of the neural network model, and 10% is used as the test set to detect the trained network model. Generalization: since the sample set used for network training contains many different parameters, it is necessary to preprocess the input data and the output after the detection results are used, so that they are all at  $[0,1]$  or  $[-1, 1]$ . In addition, we also correspondingly collected facial emotion image data of college students in class and outside class.

**4.3. Evaluation Index.** To fairly verify the performance of the proposed mental health evaluation model for college students, we use the mean square error (MSE) for evaluation, and the calculation equation is as follows:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y \wedge_i - y_i)^2. \quad (17)$$

**4.4. Experimental Results.** We first compared with the BP network. Table 2 gives the experimental results. It can be seen that the proposed model has achieved a huge performance improvement over the BP network. This proves the effectiveness of this model.

In addition, to further verify the effectiveness of the model in this article, we designed an ablation experiment. We split the dual-branch network into single branches to conduct experiments one by one. The comparative results are shown in Table 3.

From Table 2, we find that the vision-based branch is better than Bayesian-BP, which proves that the vision-based mental emotion recognition is more reliable and effective. This may be due to the fact that college students have too strong subjective awareness when filling in the SCL90 Symptom Self-Rating Scale; however, the single branch is worse than the double branch, which again illustrates the superiority of the proposed model.

We also simulated the trained network and then reverse normalized the simulation results and compared them with the original data. The comparison result of predicted data and real data is shown in Figure 8. Among them,  $O$  represents real data, and  $*$  represents simulated data. It can be seen from the figure that the obtained simulation data is very close to the

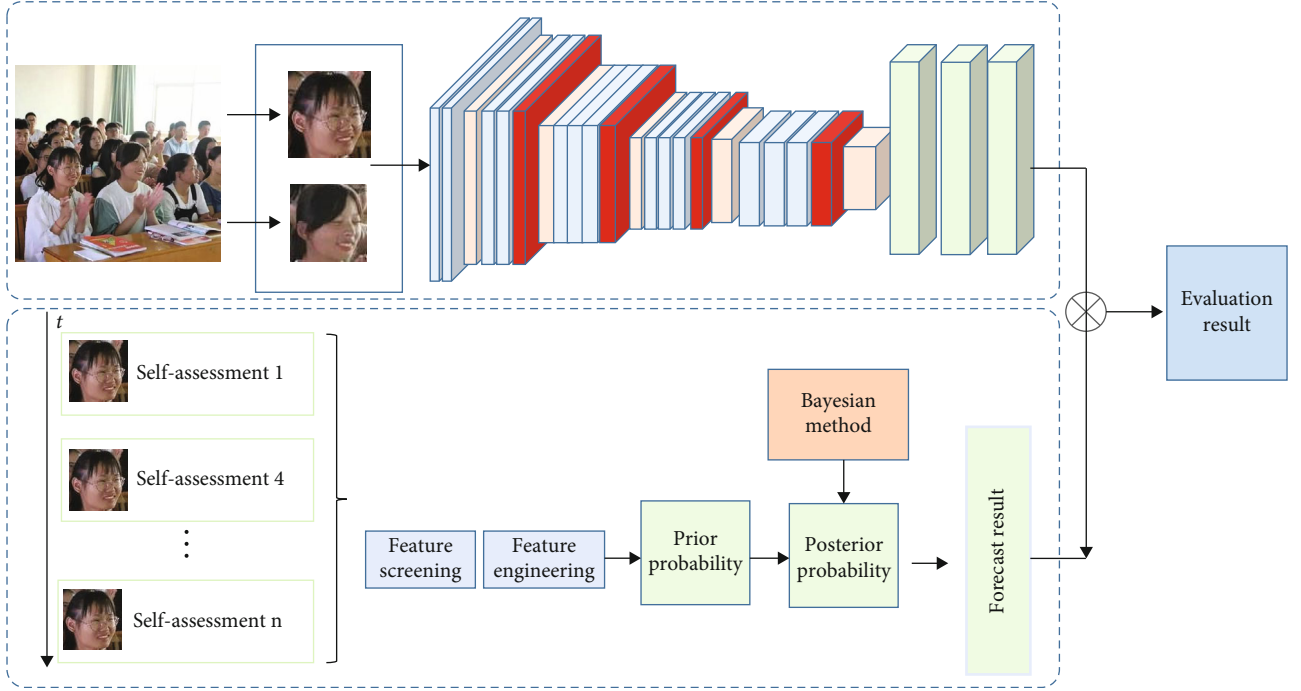


FIGURE 7: Schematic diagram of the proposed model.

TABLE 1: Environment configuration.

Item	Parameter
CPU	Intel i9 10850k
RAM	16GB
GPU	NVIDIA GTX 1080Ti
OS	Windows 10
Development tools	PyTorch4.0, CUDA8.0, cuDNN5.1, Python3.5, Python-opencv, numpy

TABLE 2: Comparative experiment results.

Method	MSE
BP	10.56%
Ours	3.12%

TABLE 3: Ablation experiment.

Method	MSE
Bayesian-BP branch	8.11%
Vision-based branch	7.65%
Ours	3.12%

real data. This shows that the trained BP network has a better fitting effect in predicting the mental state of college students.

**4.5. Ablation Experiment for VGG.** In order to prove the superiority of VGG16 in the manuscript, we have added ablation experiments in this section to investigate the perfor-

mance difference between VGG16 and VGG19 in this algorithm. It is worth noting that the ablation experiment was carried out under the same experimental conditions, and the experimental results are shown in Table 4.

It can be seen from Table 4 that after using VGG19, the MSE is relatively increased by 50.7%, because this shows that the choice of VGG16 is the best, the proposed algorithm obtains 3.12% of MSE, and VGG16 also has fewer parameters and computational complexity.

**4.6. Ablation Experiment for Attention Mechanism.** In order to verify the influence of the attention mechanism on the proposed algorithm, this section sets up an ablation experiment of the attention mechanism. No attention means not using the attention mechanism, and attention means using the attention mechanism. The results of the ablation experiment are shown in Table 5.

It can be clearly seen from Table 5 that without using the attention mechanism, the MSE rises to 7.63%, which greatly improves the error of the proposed algorithm. Therefore, this proves that the attention mechanism is effective. The mechanism improves the accuracy of facial emotion recognition.



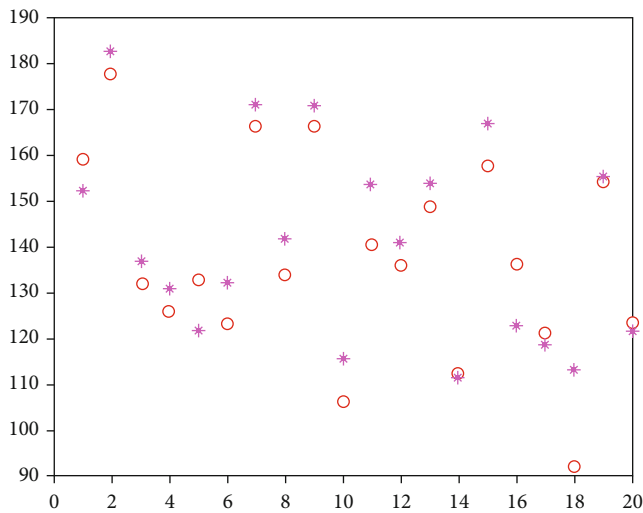


FIGURE 8: Simulation experiment results.

TABLE 4: Ablation experiments.

Method	MSE
VGG-19	5.63%
(VGG-16) ours	3.12%

TABLE 5: Ablation experiments.

Method	MSE
No attention	7.63%
(Attention) ours	3.12%

## 5. Conclusion

The mental health of college students and intervention is the most important tasks in the management of students. With the development of intelligent technology, it has become possible to evaluate and predict the mental state of college students in real time through multimodal data and neural network models. Therefore, this paper proposes a novel multimodal neural network model. Specifically, our model is divided into two branches. One branch is the traditional mental health assessment and prediction algorithm based on the improved BP neural network and the International Mental Health Scale SCL-90. Considering that this method is already difficult to meet the requirements for the accuracy of college students' mental health assessment, our other branch is the facial emotion recognition of college students based on computer vision, which is used to assist in the evaluation of mental health assessment. Through simulation and comparative experiments, we prove the effectiveness and superiority of the proposed method.

## Data Availability

The data used to support the findings of this study are included within the article.

## Conflicts of Interest

All the authors do not have any possible conflicts of interest.

## Acknowledgments

This work was supported by the 2019 Teaching Reform Research Project of Wuxi Vocational College of Science and Technology: "flipped classroom" teaching mode in college students' mental health education course application and design under Grant JG2019205.

## References

- [1] M. S. Spoorthy, S. K. Pratapa, and S. Mahant, "Mental health problems faced by healthcare workers due to the COVID-19 pandemic-a review," *Asian Journal of Psychiatry*, vol. 51, p. 102119, 2020.
- [2] R. Bruffaerts, P. Mortier, G. Kiekens et al., "Mental health problems in college freshmen: prevalence and academic functioning," *Journal of Affective Disorders*, vol. 225, pp. 97–103, 2018.
- [3] K. Levecque, F. Anseel, A. De Beuckelaer, J. Van der Heyden, and L. Gisle, "Work organization and mental health problems in PhD students," *Research Policy*, vol. 46, no. 4, pp. 868–879, 2017.
- [4] B. N. Hand, A. M. Angell, L. Harris, and L. A. Carpenter, "Prevalence of physical and mental health conditions in Medicare-enrolled, autistic older adults," *Autism*, vol. 24, no. 3, pp. 755–764, 2020.
- [5] The COVID-19 Mental Health Policy Research Unit Group, L. Sheridan Rains, S. Johnson et al., "Early impacts of the COVID-19 pandemic on mental health care and on people with mental health conditions: framework synthesis of international experiences and responses," *Social Psychiatry and Psychiatric Epidemiology*, vol. 56, no. 1, pp. 13–24, 2021.
- [6] K. Huckvale, J. Nicholas, J. Torous, and M. E. Larsen, "Smartphone apps for the treatment of mental health conditions: status and considerations," *Current Opinion in Psychology*, vol. 36, pp. 65–70, 2020.
- [7] Y. Kim, H. J. Joo, and S. Lee, "School factors related to high school dropout," *KEDI Journal of Educational Policy*, vol. 15, no. 1, 2018.
- [8] R. Samuel and K. Burger, "Negative life events, self-efficacy, and social support: risk and protective factors for school dropout intentions and dropout," *Journal of Educational Psychology*, vol. 112, no. 5, pp. 973–986, 2020.
- [9] S. McManus and D. Gunnell, "Trends in mental health, non-suicidal self-harm and suicide attempts in 16–24-year old students and non-students in England, 2000–2014," *Social Psychiatry and Psychiatric Epidemiology*, vol. 55, no. 1, pp. 125–128, 2020.
- [10] K. Russell, S. Allan, L. Beattie, J. Bohan, K. MacMahon, and S. Rasmussen, "Sleep problem, suicide and self-harm in university students: a systematic review," *Sleep Medicine Reviews*, vol. 44, pp. 58–69, 2019.
- [11] T. Curran and M. Standage, "Psychological needs and the quality of student engagement in physical education: teachers as key facilitators," *Journal of Teaching in Physical Education*, vol. 36, no. 3, pp. 262–276, 2017.

- [12] C. Valiente, J. Swanson, D. DeLay, A. M. Fraser, and J. H. Parker, "Emotion-related socialization in the classroom: considering the roles of teachers, peers, and the classroom context," *Developmental Psychology*, vol. 56, no. 3, pp. 578–594, 2020.
- [13] A. Cekaite and A. Ekström, "Emotion socialization in teacher-child interaction: teachers' responses to children's negative emotions," *Frontiers in Psychology*, vol. 10, p. 1546, 2019.
- [14] O. M. Alegre de la Rosa and L. M. Villar Angulo, "Evaluation of emotional and psycholinguistic problems in deaf and hard-of-hearing students in the Canary Islands," *Heliyon*, vol. 7, no. 3, article e06446, 2021.
- [15] C. T. Liang, J. Liu, D. Nguyen, and G. Song, "Contextualizing Asian American college student psychological health," *New Directions for Student Services*, vol. 2017, no. 160, pp. 81–92, 2017.
- [16] P. R. Selvaraj and C. S. Bhat, "Predicting the mental health of college students with psychological capital," *Journal of Mental Health*, vol. 27, no. 3, pp. 279–287, 2018.
- [17] C. D. Kouros, M. M. Pruitt, N. V. Ekas, R. Kiriaki, and M. Sunderland, "Helicopter parenting, autonomy support, and college students' mental health and well-being: the moderating role of sex and ethnicity," *Journal of Child and Family Studies*, vol. 26, no. 3, pp. 939–949, 2017.
- [18] A. Tang, R. J. Van Lieshout, A. Lahat et al., "Shyness trajectories across the first four decades predict mental health outcomes," *Journal of Abnormal Child Psychology*, vol. 45, no. 8, pp. 1621–1633, 2017.
- [19] J. Zhang, J. Sun, J. Wang, and X. G. Yue, "Visual object tracking based on residual network and cascaded correlation filters," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, pp. 8427–8440, 2021.
- [20] L. Zhang, X. Wang, X. Dong, L. Sun, W. Cai, and X. Ning, "Finger vein image enhancement based on guided tri-Gaussian filters," *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 1, pp. 17–23, 2021.
- [21] Z. Huang, P. Zhang, R. Liu, and D. Li, "Immature apple detection method based on improved Yolov3," *ASP Transactions on Internet of Things*, vol. 1, no. 1, pp. 9–13, 2021.
- [22] M. Gao, W. Cai, and R. Liu, "AGTH-Net: attention-based graph convolution-guided third-order hourglass network for sports video classification," *Journal of Healthcare Engineering*, vol. 2021, Article ID 8517161, 10 pages, 2021.
- [23] W. Cai, Z. Wei, R. Liu, Y. Zhuang, Y. Wang, and X. Ning, "Remote sensing image recognition based on multi-attention residual fusion networks," *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 1, pp. 1–8, 2021.
- [24] R. Liu, X. Ning, W. Cai, and G. Li, "Multiscale dense cross-attention mechanism with covariance pooling for hyperspectral image scene classification," *Mobile Information Systems*, vol. 2021, Article ID 9962057, 15 pages, 2021.
- [25] L. Huang, G. Xie, W. Zhao, Y. Gu, and Y. Huang, "Regional logistics demand forecasting: a BP neural network approach," *Complex & Intelligent Systems*, pp. 1–16, 2021.
- [26] X. Ning, X. Wang, S. Xu et al., "A review of research on co-training," *Concurrency and Computation: Practice and Experience*, no. article e6276, 2021.
- [27] Z. Chu, M. Hu, and X. Chen, "Robotic grasp detection using a novel two-stage approach," *ASP Transactions on Internet of Things*, vol. 1, no. 1, pp. 19–29, 2021.
- [28] B. C. Ko, "A brief review of facial emotion recognition based on visual information," *Sensors*, vol. 18, no. 2, p. 401, 2018.
- [29] D. K. Jain, P. Shamsolmoali, and P. Sehdev, "Extended deep neural network for facial emotion recognition," *Pattern Recognition Letters*, vol. 120, pp. 69–74, 2019.
- [30] D. Mehta, M. F. H. Siddiqui, and A. Y. Javaid, "Facial emotion recognition: a survey and real-world user experiences in mixed reality," *Sensors*, vol. 18, no. 2, p. 416, 2018.
- [31] B. Fasel, "Head-pose invariant facial expression recognition using convolutional neural networks," in *Proceedings. Fourth IEEE international conference on multimodal interfaces*, pp. 529–534, Pittsburgh, PA, USA, October 2002.
- [32] M. Matsugu, K. Mori, Y. Mitari, and Y. Kaneda, "Subject independent facial expression recognition with robust face detection using a convolutional neural network," *Neural Networks*, vol. 16, no. 5-6, pp. 555–559, 2003.
- [33] A. Yao, D. Cai, P. Hu, S. Wang, L. Sha, and Y. Chen, "HoloNet: towards robust emotion recognition in the wild," in *Proceedings of the 18th ACM international conference on multimodal interaction*, pp. 472–478, New York, NY, USA, October 2016.
- [34] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826, Las Vegas, NV, USA, 2016.
- [35] X. Zhao, X. Shi, and S. Zhang, "Facial expression recognition via deep learning," *IETE Technical Review*, vol. 32, no. 5, pp. 347–355, 2015.
- [36] B. Sun, L. Li, T. Zuo, Y. Chen, G. Zhou, and X. Wu, "Combining multimodal features with hierarchical classifier fusion for emotion recognition in the wild," in *Proceedings of the 16th international conference on multimodal interaction*, pp. 481–486, New York, NY, USA, November 2014.
- [37] K. Zhang, Y. Huang, Y. Du, and L. Wang, "Facial expression recognition based on deep evolutionary spatial-temporal networks," *IEEE Transactions on Image Processing*, vol. 26, no. 9, pp. 4193–4203, 2017.
- [38] W. Cai and Z. Wei, "Remote sensing image classification based on a cross-attention mechanism and graph convolution," in *IEEE Geoscience and Remote Sensing Letters*, pp. 1–5, InPress, 2020.
- [39] T. Heseltine, N. Pears, and J. Austin, "Three-dimensional face recognition using combinations of surface feature map subspace components," *Image and Vision Computing*, vol. 26, no. 3, pp. 382–396, 2008.

## Research Article

# Research on a High-Speed and Heavy-Duty Closed-Loop Drive System of a Two-Phase Hybrid Stepping Motor Based on a Hybrid Controller

Zhou Yansuo <sup>1,2</sup>, Leng Yonggang <sup>1</sup>, Lu Wenqi <sup>3</sup>, Li Yu,<sup>4</sup> Li Qingmian,<sup>2</sup> and Wu Di<sup>3</sup>

<sup>1</sup>School of Mechanical Engineering, Tianjin University, Tianjin 300072, China

<sup>2</sup>Zhejiang Institute of Industry and Information Technology, Hangzhou 310006, China

<sup>3</sup>School of Mechanical Engineering and Automation, Zhejiang Sci-Tech University, Hangzhou 310018, China

<sup>4</sup>Zhejiang Institute of Modern Agricultural Equipment Design and Research, Hangzhou 310009, China

Correspondence should be addressed to Leng Yonggang; leng\_yg@tju.edu.cn and Lu Wenqi; luwenqi@zstu.edu.cn

Received 10 May 2021; Accepted 19 July 2021; Published 26 August 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Zhou Yansuo et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Low load capacity and poor positioning accuracy of stepper motors in high-speed operation are currently two of the bottlenecks that limit their application in high-speed and heavy-duty drive applications. To solve this problem, a hybrid controller is proposed for the high-speed heavy-duty closed-loop stepper motor driving system, which includes two core contents. First, for the position control, a hybrid controller based on position error for open-loop/closed-loop automatic switching-combined spatial current given amplitude and angle automatic adjustment is proposed. Secondly, an advanced angle compensation strategy based on error-integrated feedforward is adopted to compensate for the electrical angle of the combined space current vector. To verify the effectiveness of the proposed method, theoretical analysis and system development as well as testing are carried out. Compared with the traditional open-loop drive system, results show that the maximum operating speed and maximum torque of the newly developed drive system based on the proposed method are improved by 50% and 81.25%, respectively. And at the same set speed and position, the response speed is faster and the accuracy of the steady-state process is higher. In the case of setting higher running speed and load torque, the drive system also maintains high-precision operation.

## 1. Introduction

A two-phase hybrid stepping motor is widely used in 3D printing, electric vehicles, textile machinery, industrial robots, etc. due to its simple control, accurate positioning, low cost, etc. [1–5]. But most of the current applications of the stepping motor is mainly open-loop control, which has shortcomings such as low-frequency oscillation, low speed, poor load capacity, and out of step. In recent years, with the continuous development of power electronics technology and modern industrial application requirements, higher requirements (high precision, high response, and high load capacity) for the existing open-loop control and the control performance of stepper motors are put forward. The traditional open-loop methods [4, 5] can no longer meet the driving requirements.

Some improved control strategies have also presented in recent years. A nuclear step control strategy is proposed for a two-phase hybrid stepping motor [6, 7]. Based on the traditional open-loop control of the stepper motor, the real-time rotor position feedback is introduced by detecting the nuclear step counter. Position compensation is completed by increasing or decreasing control pulses when out of step or beyond step. The accuracy of its position was ensured to some extent, but this control strategy cannot increase the output torque when the motor is out of step or blocked. So the load capacity using this strategy is weak. A power angle control strategy of the stepping motor is proposed in [8, 9] by monitoring the power angle in real time and ensuring that the power angle does not exceed the steady-state operation of the motor by adjusting the terminal voltage of the motor phase winding and the rotor speed of the stepping motor. When the power

angle is close to the maximum power angle of the stepper motor in steady-state operation, the excitation current of the motor winding is controlled to achieve the purpose of enhancing the antiload fluctuation capability of the stepping motor. The method can improve the load capacity of the motor. However, the acquirement of the stepping motor power angle is extremely complicated. So it is difficult to realize in actual occasion. Moreover, the test process finds that the motor driven by the method has a higher temperature rise and lowers the service life of the motor.

Research efforts have also been made to implement sensorless control for stepper motors [10–15]. The majority of these sensorless methods [11, 12] are based on observer techniques and Kalman filtering. The tuning of such observers for sensorless motor control is complex and time-consuming. Moreover, most observer-based sensorless techniques [11, 12] rely on models containing load parameters such as inertia and friction. Load-dependent parameters of the motor control system are difficult to determine and often vary over time. On the other hand, the rotor position based on the response of high-frequency test signals is estimated in [12]. These methods estimate the rotor position accurately, even at low speeds where observer-based methods often fail. However, these methods need access to the power electronics to generate the desired test pulses. This hinders implementation at existing stepping motor controllers. The sensorless algorithm, described in [13–15], estimates the load angle. In contrast to more complex observer algorithms, this estimator does not depend on mechanical load parameters and is characterized by a low computational cost. However, the system designed by the sensorless estimation method has poor high-speed load capacity and low positioning accuracy, which cannot meet the requirements of high-speed and high-load driving.

For this purpose, some field-oriented controllers with position feedback and vector control algorithms are used to drive stepper motors [16–24]. This method improves the positioning accuracy of the stepping motor, but compared with the general DC or AC motor, the air gap of the hybrid stepping motor is much smaller. The reluctance torque has a large influence on the electromagnetic torque and cannot be ignored. It is more complicated to obtain accurate torque expression. The closed-loop drive system designed by this scheme has a relatively low load capacity during high-speed operation. To reduce the torque ripple [18–21] or improve the maximum speed [22], some research efforts have been carried out, sometimes even using artificial intelligence techniques such as neural networks [23, 24] and fuzzy control [25], but these algorithms are more complex to implement and will be limited by hardware costs.

Therefore, the above existing research work focuses on how to improve the positioning accuracy of the stepping motor and has achieved certain improvements. Some applications (engraving machine, stripping machine, etc.) have been applied in some occasions where the running speed and load capacity are not high (generally between 300 and 600 r/min). However, in some occasions where the running speed is higher (usually between 800 and 1200 r/min) and the load capacity is strong (template machine, embroidery

machine, etc.), the current algorithm cannot meet the driving requirements, and it is urgent to introduce a control method to improve the high-speed load capacity of the motor. To this end, to improve the load capacity and positioning accuracy of the motor at high speed, the control method and system in this paper are proposed to carry out theoretical and experimental research.

## 2. The Closed-Loop Driving Method Based on the Hybrid Controller

To improve the positioning accuracy and load capacity of the stepping motor under high-speed operation, this paper proposes a high-speed heavy-duty driving method and system based on a hybrid controller. The principal block diagram is shown in Figure 1. It is composed of a position hybrid controller, a current loop PI regulator, Clark transform, a two-phase space vector modulation algorithm, an H-bridge inverter, lead angle compensation, and a hybrid stepping motor (self-contained encoder). The control scheme of the system adopts the position/current two-loop control structure. Compared with the existing position/speed/current three-loop closed-loop control structure, the following improvements are made:

- (1) Position control proposes an open loop based on position error

In position control, a hybrid controller is proposed, which can automatically switch the open loop/closed loop according to the position error and automatically adjust the given amplitude and angle of the synthetic space current. Its output is the two-phase space current vector of the synthetic current space vector in the stator coordinate system.

- (2) Propose the error-integrated feedforward-based advancement

The angle compensation strategy compensates for the electrical angle of the resultant space current vector. The two improved methods are specifically described below.

The main parameters involved in the control system are as follows:

$i_{ref}$  is the current reference vector,  $\theta_g$  is the current vector angle,  $i_{\alpha ref}$  is the  $\alpha$ -axis current reference vector,  $i_{\beta ref}$  is the  $\beta$ -axis current reference vector,  $i_{openhold}$  is the open-loop hold current,  $I_p$  is the maximum current output value of the driver,  $e_{pos}$  is the position error, and  $\theta_i$  is the integral calculation value of the position error.

**2.1. Position Hybrid Controller.** Based on the analysis of the advantages and disadvantages of the current position, speed, and current three-loop vector control scheme of the permanent magnet AC servo system and the stepping motor open-loop control scheme, a hybrid controller is proposed, which is based on position error for open-loop/closed-loop automatic switching, and the assignment of the current vector magnitude  $i_{ref}$  and angle  $\theta_g$  of the combined space is automatically adjusted according to the position error. Its principal block diagram is shown in Figure 2. The value of

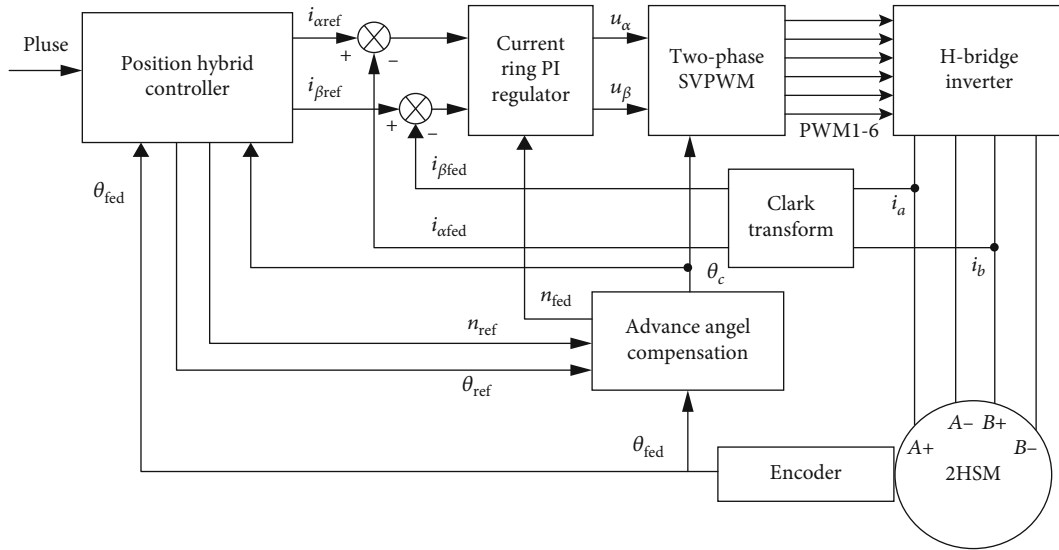


FIGURE 1: The system principle block diagram of the proposed scheme.

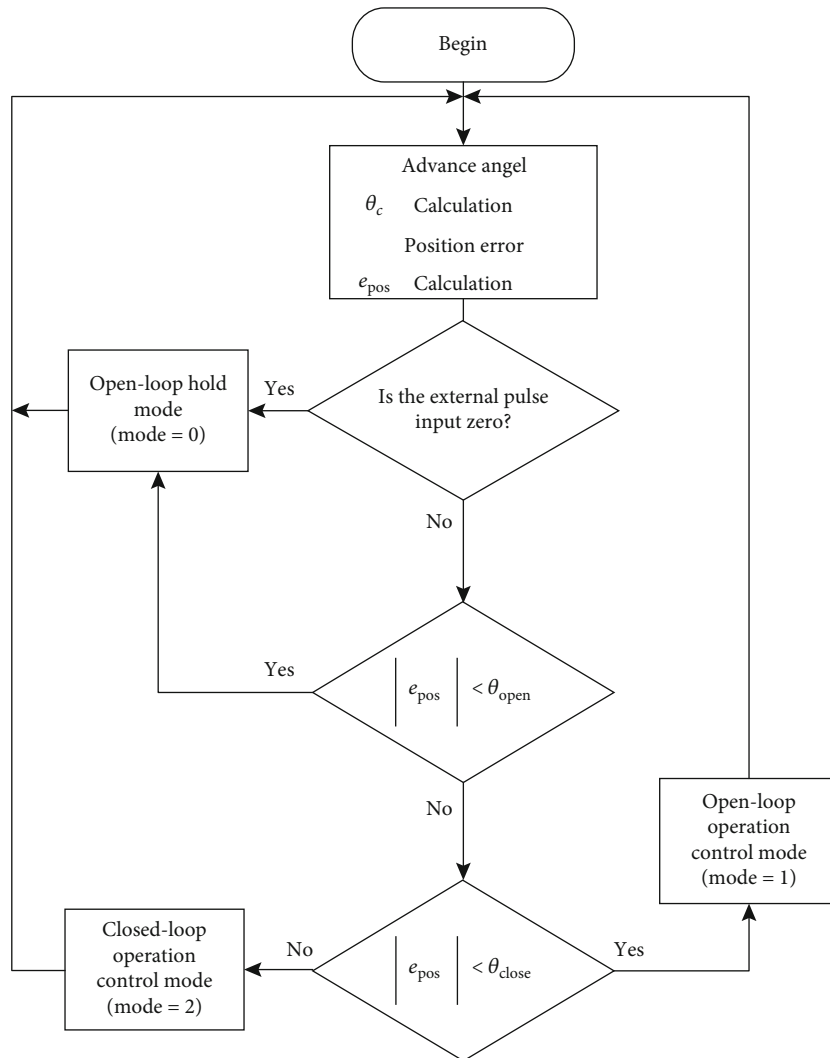
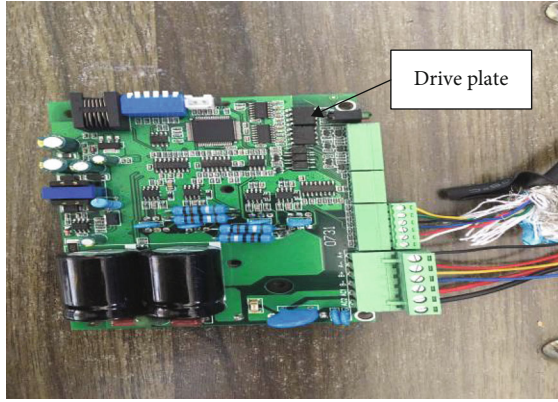
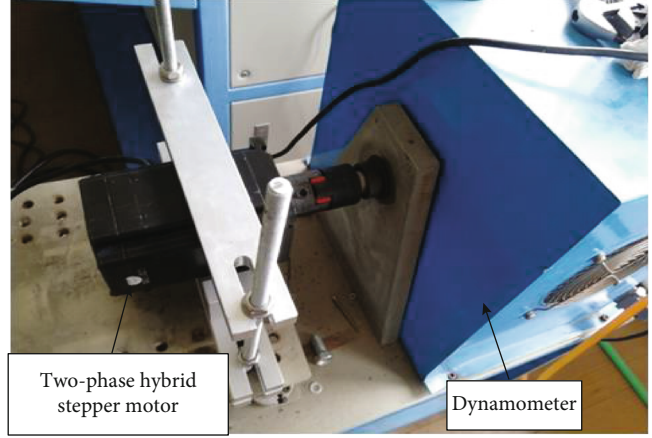


FIGURE 2: The principle block diagram of the hybrid controller based on position error automatic switching control mode.



Drive system

(a) Drive system



Stepper motor and loading equipment

(b) Stepper motor and loading equipment



Waveform acquisition system of upper computer

(c) Waveform acquisition system of the upper computer

FIGURE 3: Test platforms.

the current vector amplitude  $i_{ref}$  is adjusted according to formula (1), and the current vector angle  $\theta_g$  is adjusted according to formula (2). When  $i_{ref}$  and  $\theta_g$  are assigned, the hybrid controller can calculate the values of  $i_{\alpha ref}$  and  $i_{\beta ref}$  in the stationary coordinate system according to formula (3).

The working principle of the hybrid controller is briefly described as follows:

- (1) When the drive system does not receive the external pulse signal, the motor enters the open-loop hold mode (mode = 0). At this time, set the amplitude  $i_{ref}$  of the composite space current vector to the open-loop hold current  $i_{openhold}$ , and set the composite space current vector angle  $\theta_g$  to  $\theta_i$ , where  $i_{openhold} = P_{openhold} \times I_p$ , where  $I_p$  is the maximum current output value of the driver and  $P_{openhold}$  is the percentage of the open-loop holding current and the maximum current output of the driver.  $\theta_i$  is the integral calcula-

tion value of the position error  $e_{pos}$ , and the integral of the position error is used as a calculation to improve the positioning accuracy and output when the motor is stationary

- (2) When the drive system receives the external pulse input signal and when the absolute value of the position error is less than or equal to  $\theta_{close}$ , the motor adjusts the control mode according to the position error and the magnitude of the lead compensation angle: when the position error and the lead angle are both less than or equal to  $\theta_{close}$ , the system enters the open-loop hold mode (mode = 0). At this time, set the amplitude  $i_{ref}$  of the composite space current vector to  $i_{openhold}$  and set the angle  $\theta_g$  of the composite space current vector to  $\theta_i$ ; when the position error and the lead angle are both greater than  $\theta_{open}$  and less than or equal to  $\theta_{close}$ , it enters the open-loop operation control mode (mode = 1), and set the amplitude

TABLE 1: Parameters related to the stepper motor and drive system.

Variable name	Numerical value	Variable name	Numerical value
Polar logarithm	50	Rotor inertia	2.8 (kg·cm <sup>2</sup> )
Rated speed $n_s$	1500 r/min	Rated phase current	6 A
Hold torque	8 N·m	$P_{\text{closehold}}$	100%
Inductance	mH	$P_{\text{openhold}}$	40%
Resistance	$\Omega$	$\theta_{\text{open}}$	90 electric angle
Maximum output current of the driver	8 A	$\theta_{\text{close}}$	135 electric angle
Encoder line number	1000	$l$	1.5

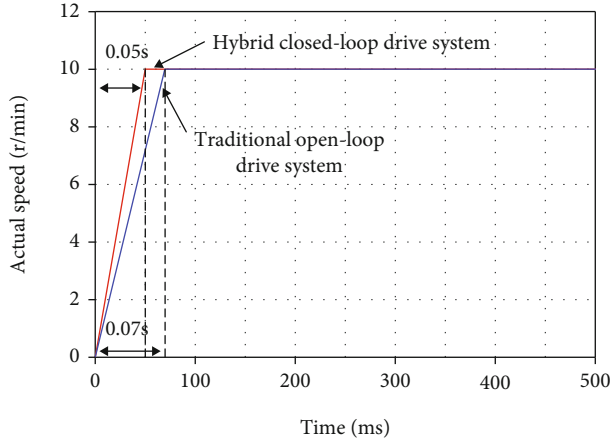


FIGURE 4: Comparison of motor speed curves at a given speed 10 r/min.

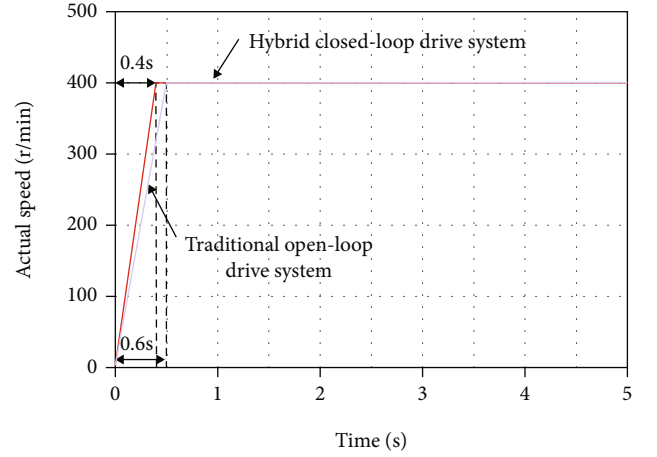


FIGURE 5: Comparison of motor speed curves at a given speed 400 r/min.

$i_{\text{ref}}$  of the combined space current vector to the open-loop running current  $i_{\text{openrun}}$ , wherein the calculation formula of  $i_{\text{openrun}}$  is shown in formula (4), which adjusts in real time according to the position error; set the angle  $\theta_g$  of the combined space current vector to the given position angle  $\theta_{\text{ref}}$ . It adjusts the positioning accuracy and updates the real-time angle according to the externally set subdivision value

- (3) When the drive system receives the external pulse input signal and when the position error is greater than  $\theta_{\text{close}}$ , the motor enters the closed-loop operation control mode (mode = 2). At this time, the amplitude  $i_{\text{ref}}$  of the combined space current vector is set as the closed-loop running current  $i_{\text{closerun}}$ ; in the setting space, the angle  $\theta_g$  of the combined current vector is set to  $\theta_{\text{fed}} + \theta_c$ , where  $i_{\text{closerun}} = P_{\text{closehold}} \times \mathbf{I_p}$ , where  $P_{\text{closehold}}$  is the percentage of the closed-loop operating current and the maximum current output of the driver,  $\theta_{\text{fed}}$  is the angular value of the actual feedback of the encoder, and  $\theta_c$  is based on the position error and speed error compensation calculation. The obtained lead angle is used to improve the motor output and positioning accuracy during high-speed operation

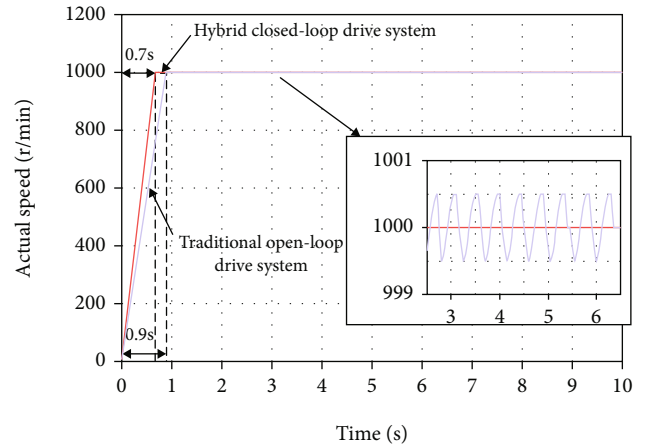


FIGURE 6: Comparison of motor speed curves at a given speed 1000 r/min.

$$i_{\text{ref}} = \begin{cases} i_{\text{openhold}}, & |e_{\text{pos}}| < \theta_{\text{open}}, \\ i_{\text{openrun}}, & \theta_{\text{open}} \leq |e_{\text{pos}}| < \theta_{\text{close}}, \\ i_{\text{closerun}}, & |e_{\text{pos}}| \geq \theta_{\text{close}}, \end{cases} \quad (1)$$

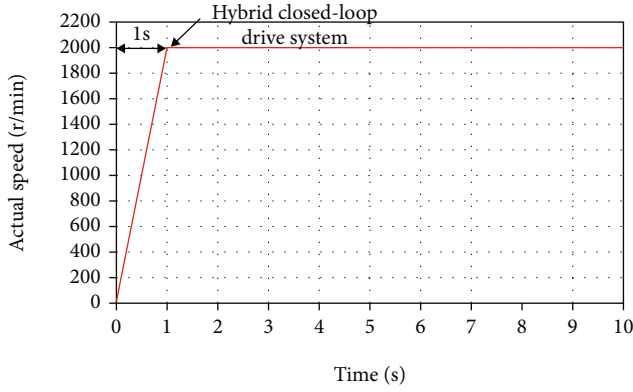


FIGURE 7: Comparison of motor speed curves at a given speed 2000 r/min.

$$\theta_g = \begin{cases} \theta_i, & \text{Mode} = 0, \\ \theta_{\text{ref}}, & \text{Mode} = 1, \\ \theta_{\text{fed}} + \theta_c, & \text{Mode} = 2, \end{cases} \quad (2)$$

$$\begin{aligned} i_{\text{aref}} &= i_{\text{ref}} \times \cos(\theta_g), \\ i_{\beta\text{ref}} &= i_{\text{ref}} \times \sin(\theta_g), \end{aligned} \quad (3)$$

$$i_{\text{openrun}} = \mathbf{I} \mathbf{p} \times \left( P_{\text{openhold}} + \frac{(P_{\text{closehold}} - P_{\text{openhold}})}{\theta_{\text{close}} - \theta_{\text{open}}} (|e_{\text{pos}}| - \theta_{\text{open}}) \right). \quad (4)$$

**2.2. Leading Angle Control Based on Position Error Integration Feedforward.** In this paper, a lead angle control method based on error-integrated feedforward compensation is proposed. The lead angle consists of two parts: position error lead angle and speed error lead angle. The calculation method is shown in formula (5), where  $Kp_p$  and  $Kp_n$  are the feedforward proportional coefficients of position error and velocity error, respectively.

$$\theta_c = Kp_p \times e_{\text{pos}} + Kp_n \times e_n. \quad (5)$$

### 3. Experimental Test

To verify the correctness and superiority of the proposed method, this paper designs the software and hardware of the driver and the corresponding test platform (as shown in Figure 3). Based on this, a closed-loop drive system based on the hybrid controller (hereinafter referred to as the hybrid closed-loop drive system) and the performance comparison test of the traditional open-loop stepping motor drive system (hereinafter referred to as the traditional open-loop drive system), the relevant parameters of the stepper motor and the drive system are shown in Table 1.

**3.1. Speed Control.** The comparison test was carried out under no load, given the speed of 10 r/min, and the obtained motor speed waveform is shown in Figure 4. By observing the waveform, the motor of the hybrid closed-loop drive system rises to a given speed for about 0.05 s, and the rotational

speed error of the steady-state process is almost 0 rpm. The motor of the traditional open-loop drive system responds slowly, rising to a given speed for about 0.07 s, but the rotational speed error of the steady-state process is also almost 0 rpm. Therefore, it is experimentally known that at the low-speed start, the closed-loop drive system is faster than the open-loop drive system for the response time of the motor, but in steady-state operation, the steady-state error of the closed-loop drive system and the open-loop drive system is both 0.

The comparison test was carried out under no load, given the speed of 400 r/min, and the obtained motor speed waveform is shown in Figure 5. By observing the waveform, the motor of the hybrid closed-loop drive system rises to a given speed for about 0.4 s, and the motor's rotational speed error is 0 rpm during steady-state operation. The motor response of the traditional open-loop drive system is slower, rising to about 0.5 s (0.6 s in Figure 5). For constant speed, the motor's rotational speed error during steady-state operation is 0 r/min. Therefore, it is experimentally known that the response time of the closed-loop drive system motor is faster than that of the open-loop drive system under no load and given speed 400 r/min. However, the steady-state error of both the closed-loop drive system and the open-loop drive system during steady-state operation is zero.

The comparison test was carried out under no load, given the speed of 1000 r/min, and the obtained motor speed waveform is shown in Figure 6. By observing the waveform, the motor of the hybrid closed-loop drive system rises to a given speed for about 0.7 s, and the motor's rotational speed error is 0 r/min during steady-state operation. The motor of the traditional open-loop drive system responds slowly, rising to a given speed for about 0.9 s, and the motor speed error during steady-state operation is  $\pm 0.5$  r/min. Therefore, it is experimentally known that when given speed is 1000 r/min with no load, the response time of the closed-loop drive system motor is faster than that of the open-loop drive system. In the steady-state process, the open-loop drive system begins to show steady-state error. The steady-state error of the closed-loop drive system is still zero.

The comparison test was carried out under no load, given the speed of 2000 r/min, and the obtained motor speed waveform is shown in Figure 7. By observing the waveform, the motor of the hybrid closed-loop drive system rises to a given speed after about 1 s, and the motor's rotational speed error is  $\pm 0.5$  r/min during steady-state operation. The conventional open-loop drive system motor starts out of step at this given speed and does not reach the given speed. Therefore, it is experimentally known that the closed-loop drive system has a higher maximum speed than the open-loop drive system at no load.

**3.2. Position Control.** According to the performance analysis under the previous speed control, this paper sets the running speed of the motor to 400 r/min and compares the test with the square wave change between 0 and 5000 pluse at the given position. The obtained test waveform is shown in Figure 8. By observing the waveform, when the actual position of the motor is 0 pluse and the given position is 5000



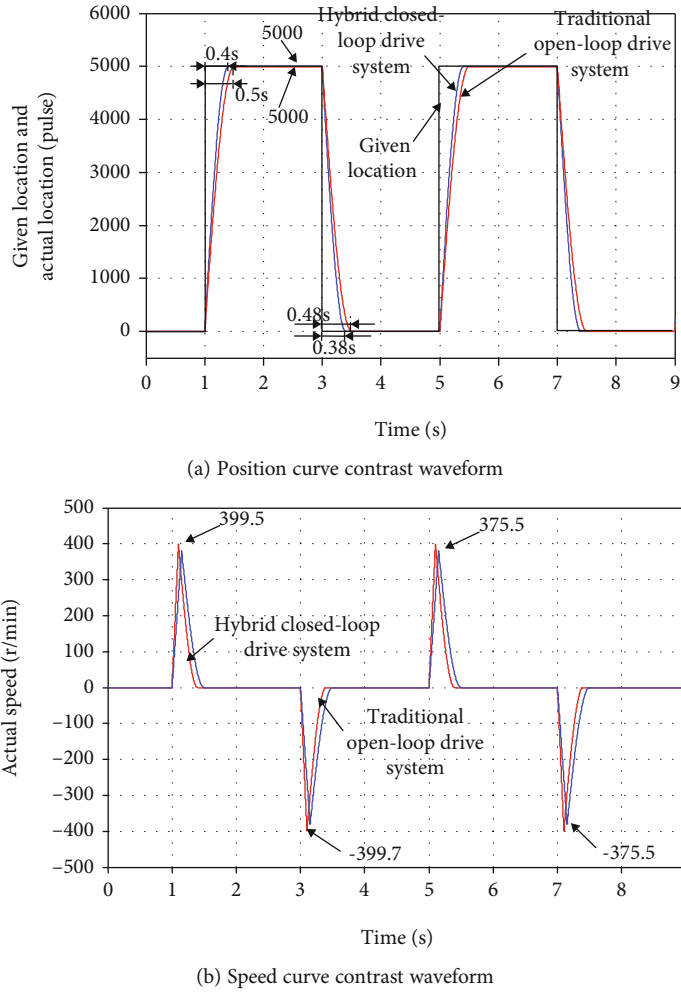


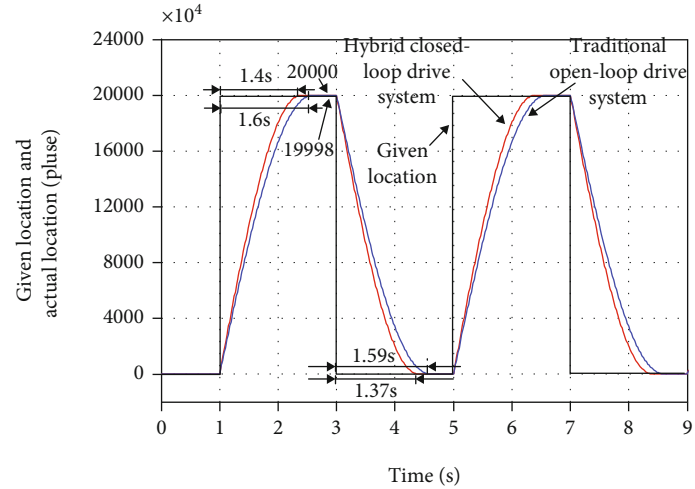
FIGURE 8: A given speed 400 r/min and a given position 5000 pulse of motor position contrast waveform.

pulse, the motor of the traditional open-loop drive system reaches the given position for about 0.5 s and the highest forward speed of the motor is 375.5 r/min, and the hybrid motor of the closed-loop drive system reaches the given position in about 0.4 s and the maximum forward speed of the motor is 399.5 r/min. In the steady-state process, the position error of both systems is 0 pulse. Therefore, from the experimental results, the positioning response performance of the closed-loop drive system is better than that of the open-loop drive system. Under the condition of no load and set running speed of 400 r/min, the positioning accuracy of the two systems is similar.

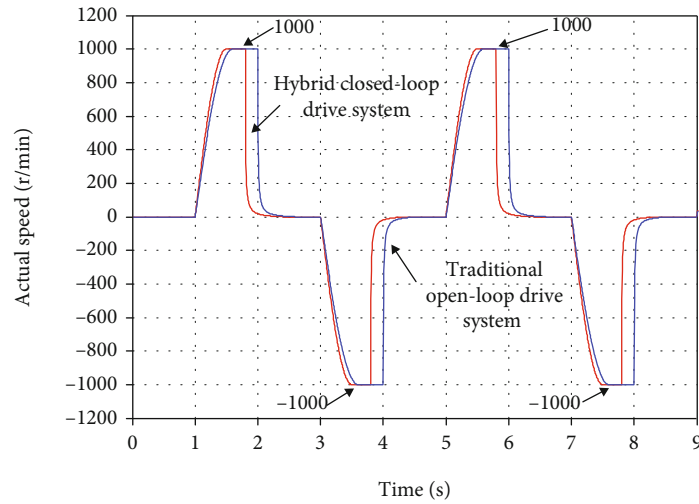
Secondly, set the running speed of the motor to 1000 r/min, and the test waveform obtained by comparing the test with the square wave change between 0 and 20000 pulse at a given position is shown in Figure 9. Observing the waveform shows that when the actual position of the motor is 0 pulse and the given position is 20000 pulse, the motor of the traditional open-loop drive system reaches the given position in about 1.6 s and the maximum forward speed of the motor is 1000 r/min. The motor of the hybrid closed-loop drive system reaches the given position in about 1.4 s. The maximum forward speed of the motor is

1000 r/min; in the steady-state process, the position error of the motor of the conventional open-loop drive system is 2 pulse; the position error of the motor of the hybrid closed-loop drive system is 0 pulse. Therefore, it is known from the experimental results that the hybrid closed-loop drive system has better positioning response performance than the traditional open-loop drive system. In the case of no load, set the running speed to 1000 r/min; the hybrid closed-loop drive system has higher positioning precision.

Also, according to the performance analysis under the previous speed control, considering the no load, the maximum operating speed of the traditional open-loop drive system is 1000 r/min, and the speed of the hybrid closed-loop drive system can be up to 2000 r/min. This paper is based on the experimental phenomena in Figure 9. Set the given speed of the hybrid closed-loop drive system motor to 2000 r/min, and also perform the comparison test under the condition that the square wave changes between 0 and 20000 pulse at a given position. The obtained test waveform is shown in Figure 10. By observing the waveform, when the actual position of the motor is 0 pulse and the given position is 20000 pulse, the motor of the hybrid closed-loop drive system reaches the given position after about 0.8 s and the



(a) Position curve contrast waveform



(b) Speed curve contrast waveform

FIGURE 9: A given speed 1000 r/min and a given position 20000 pulse of motor high-speed positioning contrast waveform.

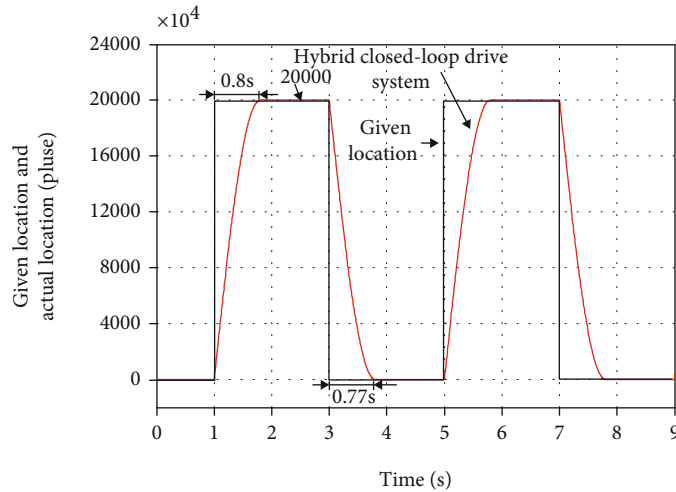
highest forward speed of the motor is 2000 r/min. In the steady-state process, the position error of the hybrid closed-loop drive system is 0 pulse. Therefore, the hybrid closed-loop drive system can further improve the positioning response speed by increasing the running speed compared to the conventional open-loop drive system when given in the same position.

**3.3. Mechanical Characteristic Curve Test.** In this paper, several sets of different speeds are set for the hybrid closed-loop drive system and the traditional open-loop drive system, respectively. The steady-state load test of the motor is performed sequentially in each set of speeds, and the maximum load torque of the system at this set of speeds is recorded. Then, through these sets of data, the mechanical characteristics of the two systems drawn are shown in Figure 11. By observing the curve, the maximum load torque of the hybrid closed-loop drive system is greater than that of the conventional open-loop drive system at the same speed. At a given speed of 1000 r/min, the maximum load torque of the hybrid

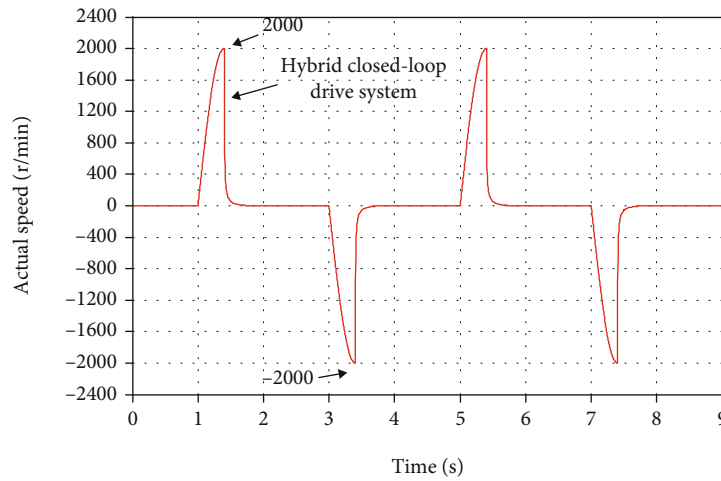
closed-loop drive system is 2 N·m, while that of the conventional open-loop drive system is 0.3 N·m. At no load, the maximum operating speed of the hybrid closed-loop drive system is 2000 r/min, while the maximum operating speed of the conventional open-loop drive system is 1000 r/min. Therefore, the traditional open-loop stepper motor drive system increases the effective load torque by 81.25% and the maximum operating speed by 50% when designed with the proposed scheme.

**3.4. On-Load Drive Performance Test.** Based on the driving performance test of the first two systems under no-load conditions and the mechanical characteristic curve of Figure 11, the driving performance test under the condition of system loading is performed below.

Firstly, in the case of a given square speed change of 400 r/min and a given position of 0-5000 pulse, the system is subjected to a comparison test of no-load and load phases; according to Figure 11, the torque corresponds to the mechanical characteristics of the traditional open-loop drive



(a) Position curve contrast waveform



(b) Speed curve contrast waveform

FIGURE 10: A given speed 2000 r/min and a given position 20000 pulse of motor high-speed positioning contrast waveform.

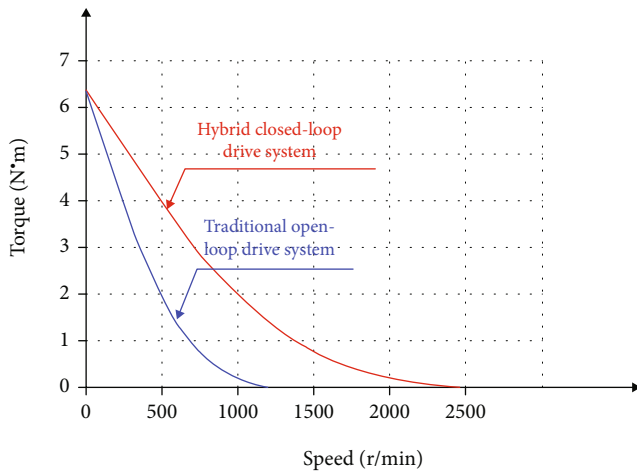
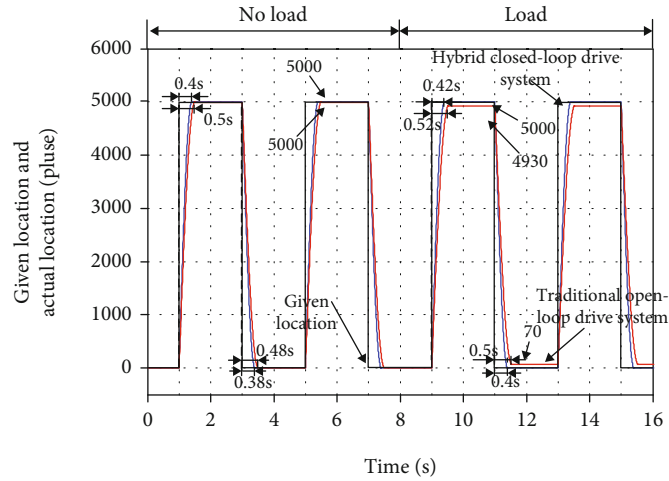


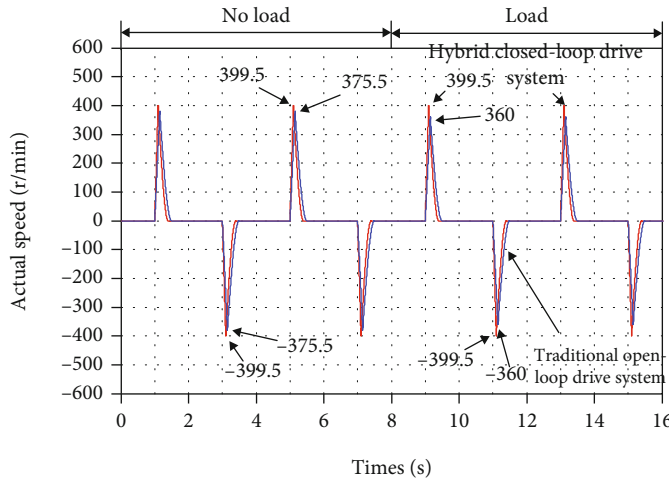
FIGURE 11: Mechanical characteristic curves of two drive systems.

system at 400 r/min. Set the load torque of the motor to 2.5 N·m, and the resulting test waveform is shown in Figure 12. By observing the waveform, it can be seen that in

the first-stage system no-load operation, the actual position of the traditional open-loop drive system rises from 0 to a given position for 0.5 s, the highest forward speed of the motor is 375.5 r/min, and the positioning error is 0 pulse. The actual position of the hybrid closed-loop drive system rises to a given position for approximately 0.4 s. The maximum forward speed of the motor is 399.5 r/min, and the positioning error is 0 pulse. In the second stage, manually loading 2.5 N·m and observing the waveform, we can see that the actual position of the traditional open-loop drive system and the given position have an error in the steady-state process (-70 pulse), which cannot be reached in the same square wave change period. At the given position, the maximum forward speed of the motor drops from 375.5 r/min to 360 r/min; the actual position of the motor of the hybrid closed-loop step drive system rises to the given position for approximately 0.42 s, and the maximum forward speed of the motor is still 399.5 r/min and the positioning error is still 0 pulse. Therefore, compared with the experimental results, the application of load reduces the speed and positioning accuracy of the traditional open-loop system, the hybrid



(a) Position curve



(b) Speed curve

FIGURE 12: Set a test with a running speed of 500 r/min, a given position of 5000 pulse, and a sudden plus 2.5 N·m load.

closed-loop drive system designed in this paper can still follow the given position well in the process of sudden loading, and the positioning accuracy has been improved.

Secondly, according to the mechanical characteristics of the hybrid closed-loop drive system in Figure 11, the corresponding torque at 400 r/min is set to 4.2 N·m for the motor; at a given speed of 400 r/min, the given position is 0. In the case of a square wave change at 5000 pulse, the system is retested in two stages of no load and loading, and the experimental phenomenon is observed. The traditional open-loop drive system has an out-of-step shutdown phenomenon, and the experimental waveform of the hybrid closed-loop drive system is shown in Figure 13; in the first stage of the system, the actual position of the hybrid closed-loop drive system rises to the given position for approximately 0.4 s, and the highest forward speed of the motor is 399.5 r/min. The error is 0 pulse; in the second stage, manually load 4.2 N·m, observe the waveform, the actual position of the hybrid closed-loop drive system rises to the given position for about 0.42 s, the highest forward speed of the motor is still 399.5 r/min, and the positioning error is still 0 pulse. Therefore, compared

with the traditional open-loop drive system, the hybrid closed-loop drive system has a higher load-carrying capacity while improving the positioning accuracy.

Under the condition that the given speed is 1000 r/min and the given position is 0-20000 pulse, the system is tested in two stages of no load and loading. According to the mechanical characteristics of the traditional open-loop drive system in Figure 11, the corresponding moment at 1000 r/min is set to be 0.3 N·m, and the test wave is obtained. It is shown in Figure 14. The observed waveforms show that in the first stage of no-load operation, the actual position of the traditional open-loop drive system rises to a given position after 1.6 s, the maximum forward speed of the motor is 1000 r/min, and the positioning error is 2 pulse; the actual position of the hybrid closed-loop drive system rises to a given position after 1.4 s, and this is the maximum positive position of the motor. The speed is 1000 r/min, and the positioning error is 0 pulse. In the second stage, 0.3 N·m is loaded manually, and the observed waveforms show that the actual position and given position of the traditional open-loop drive system have errors in the steady-state process (-190 pulse).

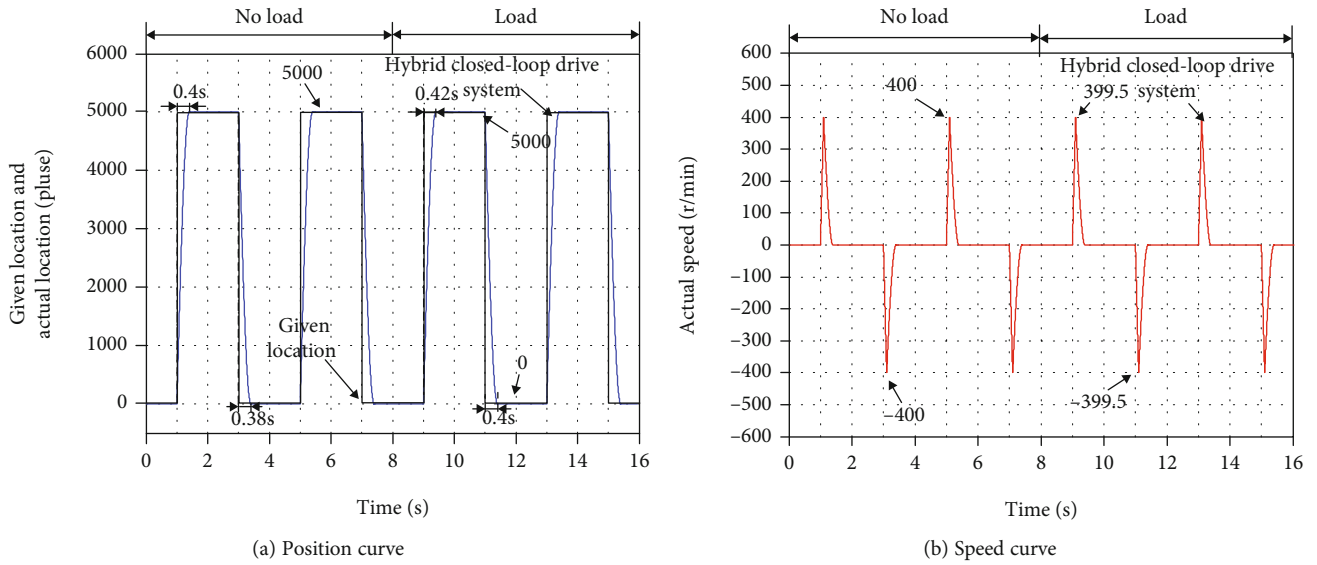


FIGURE 13: Set a test with a running speed of 500 r/min, a given position of 5000 pluse, and a sudden plus 4.2 N·m load.

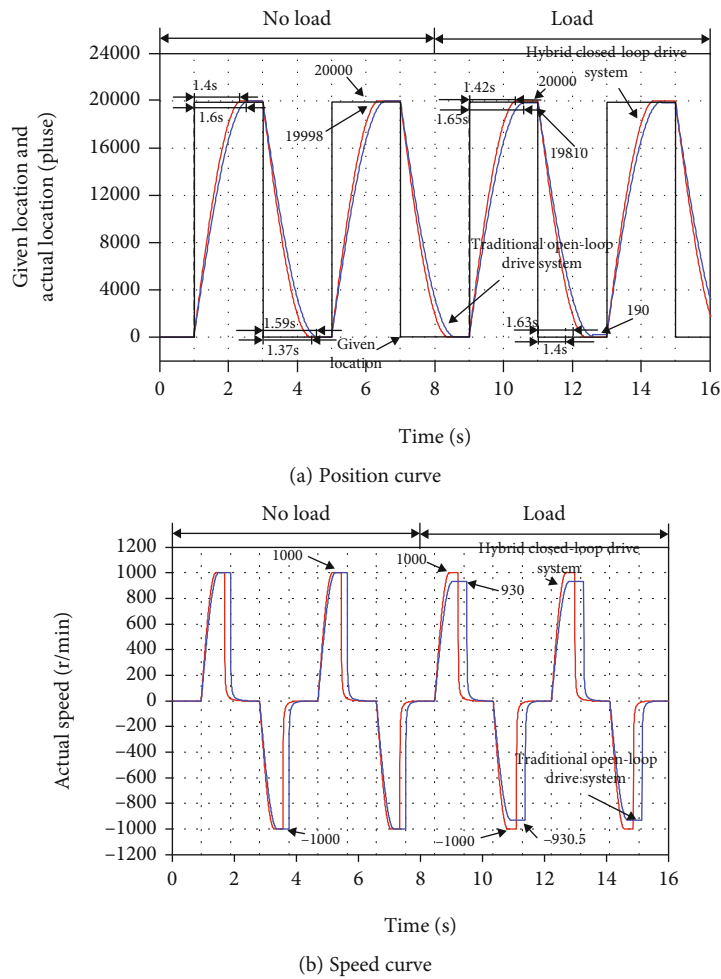


FIGURE 14: Set a test with a running speed of 1000 r/min, a given position of 20000 pluse, and a sudden plus 0.3 N·m load.

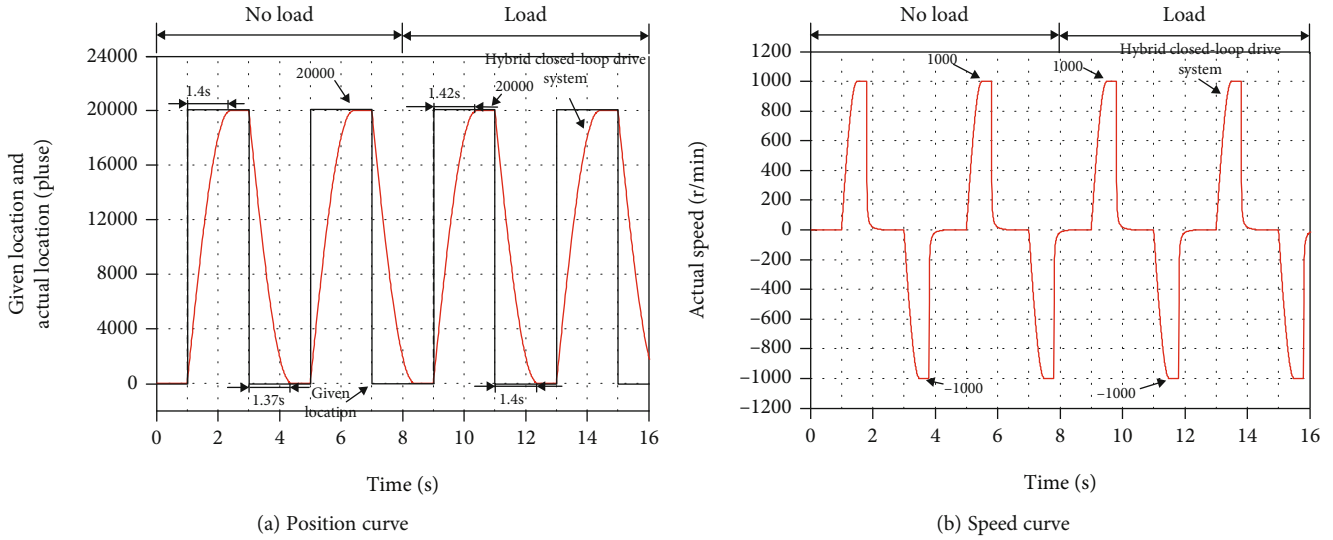


FIGURE 15: Set a test with a running speed of 1000 r/min, a given position of 20000 pulse, and a sudden plus 2 N·m load.

Within the same square wave change period, the given position cannot be reached. The maximum forward speed of the motor decreases from 1000 r/min to 930 r/min; the actual position of the hybrid closed-loop drive system rises to a given position after 1.42 s. The maximum forward speed of the motor is 1000 r/min, and the positioning error is still 0 pulse. Therefore, compared with the experimental results, the application of load reduces the speed and positioning accuracy of the traditional open-loop system, the hybrid closed-loop drive system designed in this paper can still follow the given position well in the process of sudden loading, and the positioning accuracy has been improved.

Then, according to the mechanical characteristics of the hybrid closed-loop drive system in Figure 11, with the corresponding torque at the 1000 r/min, the loading torque of the set motor is 2 N·m, in the case of a given speed of 1000 r/min and a given position of 0-20000 pulse in the square wave change. The system is compared with two stages of no load and loading. Observe the test phenomenon, the traditional open-loop drive system has an out-of-step downtime phenomenon, and the hybrid closed-loop drive system test waveform is shown in Figure 15; in the first stage of the system no-load operation, the actual position of the hybrid closed-loop drive system rises from about 1.4 s to a given position, the maximum forward rotational speed of the motor is 1000 r/min, and the positioning error in the steady-state process is 0 pulse. In the second stage, manually load 2 N·m, observe the waveform, the actual position of the hybrid closed-loop drive system rises from about 1.42 s to a given position, the maximum forward rotational speed of the motor is 1000 r/min, and the positioning error is still 0 pulse. Therefore, compared with the traditional open-loop drive system, the hybrid closed-loop drive system can improve the positioning accuracy and its carrying capacity is stronger.

#### 4. Conclusion

To improve the positioning accuracy and carrying capacity of the two-phase hybrid stepper motor at high speed of the motor, this paper presents a high-speed and high-overload drive method and system based on the hybrid controller, which contains the following two core contents:

- (1) A hybrid controller based on position error for open-loop/closed-loop automatic switching, with current amplitude and angle automatic adjustment, is proposed. (b) An advanced angle compensation strategy based on error-integrated feedforward is proposed. The corresponding test platform was set up for testing, and the system was tested to compare the following performance
- (2) In the case of no load, the maximum speed of the hybrid closed-loop servo system can reach 2000 r/min, while the maximum speed of the traditional open-loop system is only 1000 r/min, and the high speed running speed of the empty load increases by about 50%
- (3) In the case of the empty load with the same given speed and position pulse number, the response time of the mixed closed-loop stepper drive system to reach a given value is faster than that of the traditional open-loop stepper system
- (4) At the same operating speed, the carrying capacity of the hybrid closed-loop stepper drive system is greater than that of the traditional open-loop stepper system, and the effective torque can be increased to more than 81.25% of the retention torque
- (5) Compared with the traditional open-loop drive system, the hybrid closed-loop drive system exerts the same load torque at the same given pulse position,

and the hybrid closed-loop drive systems have faster response speed. In the process of high-speed motion, the hybrid closed-loop drive system can also maintain the high-torque operation, which ensures that the stepper motor has higher positioning accuracy in the working process

## Data Availability

The data used to support the findings of this study are included within the article.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported by the Key Research and Development Program of Science and Technology Department of Zhejiang Province (2021C01071 and 2021C2022).

## References

- [1] P. Siripala and Y. A. Sekercioglu, "A generalised solution for generating stepper motor speed profiles in real time," *Mechatronics*, vol. 23, no. 5, pp. 541–547, 2013.
- [2] S. Derammelaere, B. Vervisch, F. de Belie et al., "The efficiency of hybrid stepping motors: analyzing the impact of control algorithms," *IEEE Industry Applications Magazine*, vol. 20, no. 4, pp. 50–60, 2014.
- [3] Z. Dong, "Nonlinear power-level control design for MHTGRs by considering stepper motor dynamics," *Progress in Nuclear Energy*, vol. 78, pp. 216–230, 2015.
- [4] C. Gradl, A. Plöckinger, and R. Scheidl, "Sensorless position control with a hydraulic stepper drive – concept, compression modeling and experimental investigation," *Mechatronics*, vol. 35, pp. 91–101, 2016.
- [5] A. Bellini, C. Concari, G. Franceschini, and A. Toscani, "Mixed-mode PWM for high-performance stepping motors," *IEEE Transactions on Industrial Electronics*, vol. 54, no. 6, pp. 3167–3177, 2007.
- [6] J. Pillans, "Reducing position errors by vibration optimization of stepper motor drive waveforms," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 6, pp. 5176–5183, 2021.
- [7] W. Kim, Y. Lee, D. Shin, and C. C. Chung, "Nonlinear gain position control using only position feedback for permanent magnet stepper motors," *IEEE Transactions on Power Electronics*, vol. 36, no. 7, pp. 8506–8516, 2021.
- [8] F. Wang, S. Yuan, and P. Jing, "Research on the fuzzy-based out-of-step preventive control for low-speed permanent magnet linear synchronous motor," in *IEEE ICCA 2010*, pp. 1025–1029, Xiamen, China, 2010.
- [9] X. Wang, S. Lu, and S. Zhang, "Rotating angle estimation for hybrid stepper motors with application to bearing fault diagnosis," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 8, pp. 5556–5568, 2020.
- [10] S. Moon and D. H. Kim, "Step-out detection and error compensation for a micro-stepper motor using current feedback," *Mechatronics*, vol. 24, no. 3, pp. 265–273, 2014.
- [11] A. Masi, M. Butcher, M. Martino, and R. Picatoste, "An application of the extended Kalman filter for a sensorless stepper motor drive working with long cables," *IEEE Transactions on Industrial Electronics*, vol. 59, no. 11, pp. 4217–4225, 2012.
- [12] M. Bendjedja, Y. Ait-Amirat, B. Walther, and A. Berthon, "Position control of a sensorless stepper motor," *IEEE Transactions on Power Electronics*, vol. 27, no. 2, pp. 578–587, 2012.
- [13] S. Derammelaere, C. Debruyne, F. De Belie, K. Stockman, and L. Vandeveldel, "Load angle estimation for two-phase hybrid stepping motors," *IET Electric Power Applications*, vol. 8, no. 7, pp. 257–266, 2014.
- [14] S. Derammelaere, B. Vervisch, J. De Viaene, and K. Stockman, "Sensorless load angle control for two-phase hybrid stepper motors," *Mechatronics*, vol. 43, no. 3, pp. 6–17, 2017.
- [15] S. Derammelaere, F. Verbelen, and K. Stockman, "Robust sensorless load angle control for stepping motors," in *2015 18th International Conference on Electrical Machines and Systems (ICEMS)*, Pattaya, Thailand, 2015.
- [16] W. Kim, C. Yang, and C. C. Chung, "Design and implementation of simple field-oriented control for permanent magnet stepper motors without DQ transformation," *IEEE Transactions on Magnetics*, vol. 47, no. 10, pp. 4231–4234, 2011.
- [17] K. Wonhee, S. Donghoon, and C. C. Chung, "Microstepping with nonlinear torque modulation for permanent magnet stepper motors," *IEEE Transactions on Control Systems Technology*, vol. 21, no. 5, pp. 1971–1979, 2013.
- [18] W. Kim, D. Shin, Y. Lee, and C. C. Chung, "Simplified torque modulated microstepping for position control of permanent magnet stepper motors," *Mechatronics*, vol. 35, pp. 162–172, 2016.
- [19] S.-K. Kim and C. K. Ahn, "Position regulator with variable cut-off frequency mechanism for hybrid-type stepper motors," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 67, no. 10, pp. 3533–3540, 2020.
- [20] S. Rustemli, M. Yilmaz, and M. Demirtas, "Ripple reduction at speed and torque of step motors used on a two-axis robot arm," *Robotics and Computer-Integrated Manufacturing*, vol. 26, no. 6, pp. 759–767, 2010.
- [21] S. E. Lyshevski, "Microstepping and high-performance control of permanent-magnet stepper motors," *Energy Conversion and Management*, vol. 85, pp. 245–253, 2014.
- [22] A. Arias, J. Caum, and R. Griñó, "Moving towards the maximum speed in stepping motors by means of enlarging the bandwidth of the current controller," *Mechatronics*, vol. 40, no. 6, pp. 51–62, 2016.
- [23] H. M. Hasanien, "FPGA implementation of adaptive ANN controller for speed regulation of permanent magnet stepper motor drives," *Energy Conversion and Management*, vol. 52, no. 2, pp. 1252–1257, 2011.
- [24] Q. N. Le and J.-W. Jeon, "Neural-network-based low-speed-damping controller for a stepper motor with an FPGA," *IEEE Transactions on Industrial Electronics*, vol. 57, no. 9, pp. 3167–3180, 2010.
- [25] X. Liu, S. Zhen, H. Zhao, H. Sun, and Y. Chen, "Fuzzy-set theory based optimal robust design for position tracking control of permanent magnet linear motor," *IEEE Access*, vol. 7, pp. 153829–153841, 2019.

## Research Article

# A Study on the Application of Interactive English-Teaching Mode under Complex Data Analysis

Dongyang Xu <sup>1</sup> and Sang-Bing Tsai <sup>2</sup>

<sup>1</sup>Zhengzhou University of Industrial Technology, Henan 451150, China

<sup>2</sup>Regional Green Economy Development Research Center, School of Business, Wuyi University, China

Correspondence should be addressed to Dongyang Xu; toyo0228@163.com and Sang-Bing Tsai; sangbing@hotmail.com

Received 3 July 2021; Revised 2 August 2021; Accepted 9 August 2021; Published 20 August 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Dongyang Xu and Sang-Bing Tsai. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This research takes vocabulary learning in college English courses as an entry point and investigates the interrelationship between college students and electronic media, the presentation of interactive English-teaching content in college English, and the effect of different English-teaching modes on the effect of vocabulary learning through empirical methods in the form of network multimedia. This study provides ideas for an in-depth understanding of the incongruities that exist between interactive English teaching and environment, teacher and environment, and students and environment in the network multimedia environment and also brings thoughts on how to adapt interactive English teaching in college English to the development of modern information technology on the concept of interactive English teaching, interactive English-teaching methods, and interactive English-teaching modes. The concepts related to the adaptation of interactive English teaching based on multimedia, the compilation of the scale of adaptation of interactive English teaching based on multimedia and its influencing factors, and the construction of the model of influencing factors of adaptation of interactive English teaching based on multimedia, to a certain extent, enrich the theoretical system of learning adaptation. The design of the intervention model of adaptation of interactive English teaching based on multimedia broadens the intervention theory of learning adaptation; the study is also aimed at developing a multimedia-based interactive English-teaching adaptation intervention model and broadening the theory and method of learning adaptation.

## 1. Introduction

The rapid development of computer network technology has improved the efficiency and dynamism of work and learning and at the same time led to the reorganization of the education system and the redistribution of social education resources. Network technology will fundamentally improve the level of educational productivity, realize the integration and sharing of English-teaching resources, make way for diversified and open education, integrate the resources of education, and modernize the management of English teaching [1]. In terms of education methods, the Internet provides an interactive platform for information resource sharing, multidirectional mutual communication, and cooperative learning [2]. Through the promotion of the interactive platform, the inter-

active English-teaching method in the multimedia environment has changed from single and closed to diverse and open, changing the traditional English-teaching relationship between teachers and learners in different degrees, with learners taking the main identity of learning and developing the ability of independent construction. The network is a virtual platform for information transmission, reception, and sharing. Through it, the information of various points, areas, and bodies is connected to realize the sharing of these resources. It is a tool for people to exchange and use information. As a tool, it will become easier to use, and its functions will also increase. The content will become increasingly abundant. The network will use software tools such as text reading, picture viewing, video playback, download transmission, and game chat to bring people extremely rich and beautiful use



and enjoyment in terms of text, pictures, sound, and video. But it must be a tool of humankind. We believe that one day, the Internet will use the function of software tools to bring people extreme beauty and even surpass the feelings that the human body can bring.

In terms of the diversity and experiential nature of learning situations, the Web provides an environment related to “virtual reality” and multimedia technologies [3]. Through these technological environments, learners are “able to imitate and interpret the simulated world.” In this virtual environment, learners can experience the whole process of knowledge acquisition instead of just getting facts and conclusions, and the interactive English-teaching model in the multimedia environment makes the English education environment more social and enhances learners’ interest and motivation in learning [4]. At the same time, the development of the Internet has also brought about changes in the way of learning, liberating students from the traditional way of learning in which the teacher is the main body and the learners passively receive knowledge, making the way of learning diversified.

The learners can get to know diverse learning resources through the network, and on the other hand, the network provides learners with the opportunity to discover suitable learning paths for themselves and to choose suitable learning resources for learning independently [5]. Interactive English classroom English teaching in a multimedia environment is a new form of English teaching resulting from the integration of network technology and traditional classroom English teaching, with rich English-teaching resources, a multimedia form of English-teaching content, and good interactivity, which can break through the constraints of time, space, and human factors to make up for the shortcomings of traditional classroom English teaching [6]. The network-based multimedia computer-assisted English teaching provides learners with an authentic language environment, thus strengthening the learning effect and improving learners’ ability of independent and collaborative learning [7].

An interactive English classroom in the multimedia environment has certain similarities with a traditional English classroom but is significantly different from the traditional English classroom. Interactive English classroom English teaching in a multimedia environment is an English-teaching tool used to accomplish English-teaching objectives by directly intervening in the English-teaching process with the computer as the main English-teaching media in the network environment. As shown in Figure 1, it integrates text, sound, graphics, images, animation, and video, with high density, high capacity, relatively strong intuitiveness, and interest, and is not limited by time and space and can be multifunctional with all-around real virtual situations, providing an interactive space for English teaching and learning, as well as providing personalized learning services to learners.

## 2. Related Work

With the further acceleration of the process of education informatization, attention is beginning to be paid to new interactive learning methods. Among them, multimedia, as a new online course form that has emerged in recent years,

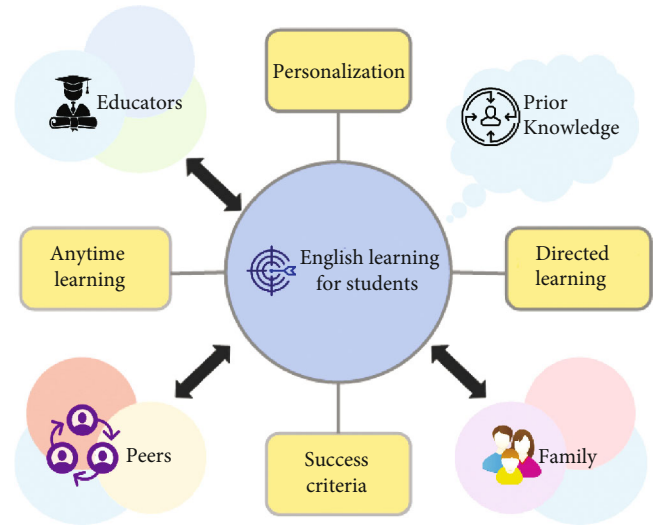


FIGURE 1: A multimedia-based model of an adaptive intervention for teaching English in an interactive English classroom.

is the latest achievement in the development of education informatization and provides new ideas and methods for the development of interactive learning in colleges and universities. At present, the use of multimedia courses for online and offline interactive learning is widely used in English teaching in colleges and universities and has been widely adopted and recognized by many schools and teachers. Therefore, the laws of teaching and learning in multimedia-based interactive learning environments have gradually become the focus of scholars’ attention to promote further reform of English language teaching in higher education through the design of scientific English-teaching models and learning activities. Ma et al. investigated a blended learning model that integrates multimedia into the traditional classroom [8]. It analyzed the benefits of integrating multimedia with classroom ELT, as well as the barriers to successful multimedia implementation. Qiao and Wang applied multimedia resources to traditional ELT for interactive learning and designed corresponding learning activities [9]. The three components include online student self-paced learning, offline focused collaborative learning, and online deepening learning. Based on the Coursera English-teaching platform, it designed a multimedia-based interactive learning model by combining face-to-face classroom English teaching and multimedia-based online learning and applied it in the English-teaching practice. Li designed a multimedia-based interactive English-teaching model, in which the learning activities on the online multimedia platform include watching videos and participating in discussions, and the offline face-to-face learning activities include teacher-student interaction and discussion and question-answer [10]. On this basis, he proposed a flow of interactive multimedia-based learning activities, including three parts: online student-oriented multimedia learning, offline teacher-led classroom English teaching, and online review after class.

Learning motivation has been a hot topic in interactive learning research, with research topics focusing on the measurement of learning motivation, factors influencing

learning motivation, the impact of learning motivation on learning outcomes, and the role of interactive learning in influencing student motivation [11]. For example, the results of Yagci's study showed that the support of course tools had a significant positive effect on learning motivation. Ling et al. analyzed the effect of learning motivation on students' learning outcomes [12]. Ma and Tsai's study also showed that learning motivation has a significant impact on students' learning effectiveness [13]. Miguel found through his study that students involved in interactive learning have higher motivation compared to face-to-face learning. Research on self-efficacy has been conducted on the effects of self-efficacy on students' learning effectiveness and the effects of interactive learning on self-efficacy. For example, Mo et al.'s study showed that multimedia-based interactive learning was effective in increasing ethnic precollege students' self-efficacy in English learning [14]. Abdelraheem's study also showed that the interactive learning approach increased students' learning efficacy. Xu and Ma studied the effect of online self-efficacy on students' learning satisfaction [15]. Ivanova's study found that students' self-efficacy was a significant predictor of learning strategy use and learning performance [16]. In the study of the psychology of learning in interactive learning environments, learning adaptation is an important research component that emphasizes the adjustment of students' attitudes, emotions, behaviors, and other elements in the learning process, which involves many psychological and behavioral elements.

The interrelationship between information technology and the degree of awareness of this and learning in the college student population is a topic that deserves attention and research. English teaching is a complex activity involving pedagogy, psychology, communication, etc. [17, 18]. The integration of network multimedia technology has brought about great changes in teaching and learning modes and means and made English teaching more complex [19]. Based on the author's many years of experience in teaching and practicing multimedia English in college English, this study focuses on each element of classroom English teaching and the relationship between them and adopts a systematic approach to the process of teaching English in college English classrooms, especially from a holistic perspective to demonstrate the use of modern information technology in college English teaching and the problems that arise in college English-teaching reform, and we analyze the causes, clarify the dysfunctions, and propose optimization strategies [20]. The study also uses various linguistic theories, English-teaching theories, information-processing theories, and psychological theories to demonstrate the interrelationship between contemporary university students and electronic media, the influence of English-teaching presentation on vocabulary learning in the form of online multimedia, and the influence of university English-teaching modes and learning styles on vocabulary learning [21]. The scope of this study covers many aspects of English teaching in the classroom, and the research approach is distinctly comprehensive. This study can provide useful guidance and reference for teachers to change the concept of English teaching, clarify the dysfunctional phenomenon, improve the way of teaching English as a foreign language in college, and enhance the effectiveness of English teaching.

### **3. Evaluation of a Model of Factors Influencing the Adaptability of Interactive English Teaching Based on Multimedia Environment**

*3.1. The Design of the Model of Factors Influencing the Adaptability of Interactive English Teaching Based on Multimedia Environment.* The theoretical basis of online teaching is the constructivist learning theory. The constructivist learning theory believes that students are the main body of the learning process and the active constructors of meaning. The acquisition of knowledge is not obtained through teaching, but learners obtain through active meaning construction under certain circumstances. One of the main theoretical foundations of traditional teaching is Ausubó's "learning and teaching" theory. Ausubó believes that teaching theory should consider problems in teaching practice. The content of the "learning and teaching" theory he put forward mainly involves three aspects: "meaningful acceptance of learning" theory, "advanced organizer" teaching strategy, and "motivation" theory. The main factors of the network teaching model are teachers, students, teaching content, teaching media, teaching, and learning support platform. In this teaching mode, teachers select teaching content according to the requirements of the syllabus, the characteristics of the students, and the cognitive foundation of the students; then design and compile them into multimedia teaching software, network courseware, or network courses; and publish them to the curriculum center. Students connect to the course center through the computer network, register and log in, and choose courses to study.

The components of learning adaptability in a general interactive English-teaching environment mainly include four dimensions: learning attitude, independent learning ability, learning environment, and physical and mental health. According to the characteristics of multimedia, which emphasizes communication and interaction, and the richness and diversity of learning resources and learning activities, this study adds the dimension of "learning communication" and appropriately revises the connotation and composition of the other dimensions mentioned above. Based on the literature, this study formed a five-part structure of multimedia-based interactive English-teaching adaptability consisting of learning attitude, independent learning ability, learning communication, learning environment, and physical and mental health. The connotation and composition of each dimension are analyzed below. The multimedia-based interactive English-teaching adaptation scale and the learning adaptation influencing factor scale were reconstituted into a new scale and distributed to students. The study first tested the overall and individual dimensional levels of college students' multimedia-based interactive English-teaching adaptation utilizing descriptive statistics and used parametric tests to analyze the variability of these data in terms of individual variables and to understand the basic situation of college students' multimedia-based interactive English-teaching adaptation. After that, we explored the logical relationships among the influencing factors of learning adaptation and the effects of each influencing factor on learning adaptation and constructed a model of influencing factors of learning adaptation.

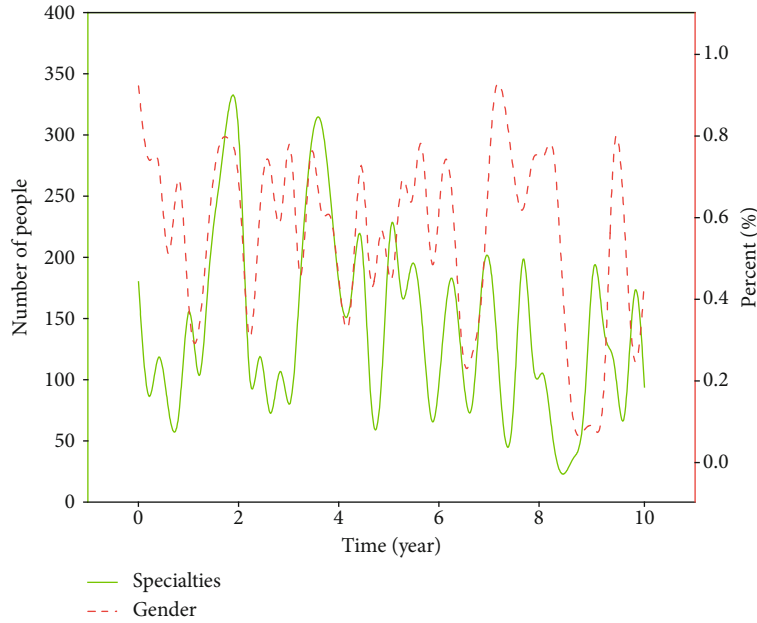


FIGURE 2: Statistics of basic information of survey respondents on the factors influencing the adaptability of interactive English teaching based on multimedia.

In terms of data information, a total of 852 questionnaires were distributed, including 483 paper questionnaires and 369 electronic questionnaires. A total of 836 questionnaires were collected, and after removing the invalid questionnaires with short answer time, missing options, choosing the same option in the whole questionnaire, and those with regular choices, 628 valid questionnaires were left, with an efficiency rate of 7.12%. In terms of basic information, it mainly includes gender, major, grade, Internet age, time of exposure to multimedia, and frequency of learning courses every month. The specific statistical results are shown in Figure 2.

Among the students who participated in this survey, in terms of gender, 22.6% of the total number are male students and 77.4% of the total number are female students. In terms of majors, 56.6% of the total number of students were in the field of literature and history, 36.5% in the field of science and technology, and 7.9% in the field of art. In terms of grades, first-year students accounted for 12.1% of the total number of students, sophomores accounted for 66.7% of the total number of students, juniors accounted for 18.0% of the total number of students, and seniors accounted for 3.2% of the total number of students. In terms of the Internet age, those with 1 year or less accounted for 2.7% of the total, those with 1-2 years accounted for 10.2%, those with 2-5 years accounted for 28.5%, and those with more than 5 years accounted for 58.6% of the total. Regarding the duration of multimedia exposure, 54.4% of the total number of students had been exposed to multimedia for 6 months or less, 13.1% of the total number of students had been exposed to multimedia for 6-12 months, 24.5% of the total number of students had been exposed to multimedia for 1-2 years, and 8.0% of the total number of students had been exposed to multimedia for more than 2 years. In terms of the frequency of courses per month, 21.7% of the total number of students attended 4 times or less, 33.4% of the total number

of students attended 4-8 times, 29.5% of the total number of students attended 8-12 times, and 15.4% of the total number of students attended more than 12 times.

From the point of view of teaching methods, every student has the opportunity and freedom to do their best and show their own strengths. This has changed the boring passive classroom teaching in the past. Every student can actively participate in learning. The enthusiasm for learning is high, which embodies the teaching principle of putting people first and focusing on the development of students and interprets the teaching concept of realizing useful mathematics for everyone, and each person has a different development in mathematics. Teaching students in accordance with their aptitude, so that every student can make progress, is my greatest wish.

This section focuses on the overall as well as individual dimensional levels of college students' adaptation to multimedia-based interactive English teaching utilizing descriptive statistics analyzes the differences in individual variables of these data utilizing parametric tests to understand the basic situation of college students' adaptation to multimedia-based interactive English teaching. In statistical analysis, the mean value is generally used to measure the overall level of the variables. Utilizing descriptive statistics, the means and standard deviations of college students' adaptation to multimedia-based interactive English teaching overall as well as for each dimension were derived, as shown in Figure 3.

The standard deviations of the dimensions as well as the overall data were within acceptable limits, i.e., the dispersion of each variable was reasonable. In terms of means, the overall level of college students' adaptability to multimedia-based interactive English language teaching was moderate ( $M = 3.56$ ), and the means of the six dimensions, in descending order, were the learning environment ( $M = 3.70$ ), learning attitude ( $M = 3.60$ ),

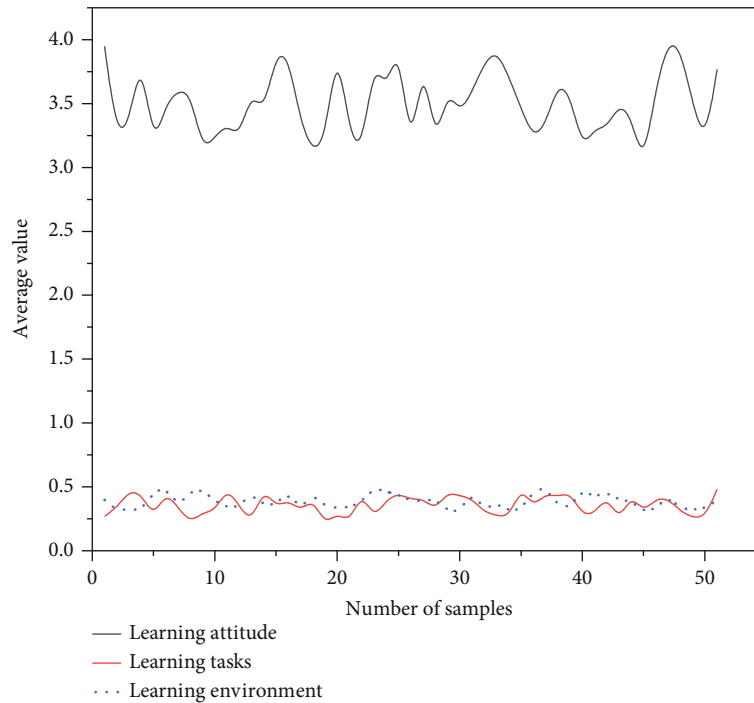


FIGURE 3: Multimedia-based interactive ELT adaptations and descriptive statistics of each dimension.

physical and mental health ( $M = 3.58$ ), learning tasks ( $M = 3.57$ ), independent learning ability ( $M = 3.47$ ), and learning communication ( $M = 3.44$ ). The results show that on specific dimensions, college students have lower scores in learning communication, independent learning ability, and learning tasks. In terms of gender, the independent sample  $t$ -test showed that the significant  $p$  values for college students' adaptation to multimedia-based interactive English language teaching and other dimensions, except for learning communication, were greater than 0.05, indicating no significant difference in gender. Specifically, in the learning communication dimension, male students were significantly higher than female students.

In terms of specialization, the one-way ANOVA showed that there were no significant differences in any of the dimensions except for the independent learning ability dimension and at the overall level. The Internet not only brings many changes to learning but also brings many changes to human life. First, it is reflected in the aspect of information transmission, then in the aspect of communication, and includes life, production, and learning. The significant  $p$  values were 0.158, 0.246, 0.519, 0.405, 0.379, and 0.725, respectively. On the independent learning ability dimension, a post hoc comparison by the LSD method based on homogeneity of the variance test revealed that the independent learning ability scores of literature and history students were higher than those of art students ( $p = 0.002$ ).

**3.2. The Construction of a Model of Factors Influencing the Adaptability of Interactive English Teaching Based on Multimedia Environment.** Based on the previous analysis, the factors influencing the adaptability of multimedia-based interactive English teaching identified in this study include learning motivation, learning self-efficacy, teacher

English teaching, learning support, multimedia platform, and course content quality. Learning self-efficacy, as an important factor influencing individual learning effort, learning persistence, and self-confidence, is the key to determining the quality of students' learning [22]. Among them, learning self-efficacy is closely related to learning motivation. For example, relevant studies have shown that learning self-efficacy is an important predictor of students' learning motivation in the online environment, and students with high self-efficacy have higher learning interests and learning ambitions, which lead to better academic performance. The self-efficacy of college students has a significant positive effect on learning motivation. Students' self-efficacy significantly affects learning motivation. The previous review of the current state of research also indicates that learning self-efficacy is a key factor influencing learning adaptability. The use of computers and other multimedia for teaching is the top priority of the content of curriculum reform. Teachers can use the multiview features of computer audiovisual media to integrate teaching content and mobilize students' learning enthusiasm.

Learning support, as a set of services that guide and assist students in their independent learning, is key to determining learning outcomes [23]. Most studies have shown that good learning support can increase students' interest in learning to motivate students to study hard. It showed that learning support provided by instructors in an online environment can significantly affect college students' motivation to learn. It was also found that by providing learning support to students, teachers can stimulate students' interest and motivation to learn and improve learning outcomes [24]. A study found that faculty support had a significant impact on students' perceived importance of the knowledge

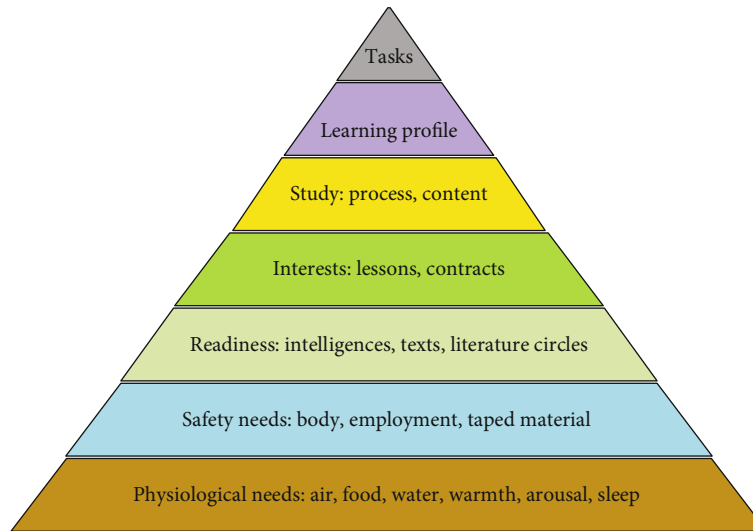


FIGURE 4: Hypothetical model of the impact of multimedia-based interactive English-teaching adaptability causal factors.

they acquired for their personal career development. Good learning support services help motivate students to learn, which in turn improves the quality of learning and thus the learning outcomes. As shown in Figure 4, the previous review of the current state of research also indicates that learning support is a key factor influencing students' learning adaptability.

The model is generally corrected according to the MI value, which indicates the minimum cardinality that the overall model will be reduced by increasing the path, i.e., the degree of improvement to the overall fitness of the model [25–27]. Therefore, attention needs to be paid to the paths with larger MI values. The principles of the correction of the model in this study include the following: the corrected index is greater than 10, the new path should have a theoretical basis, only one path is added at a time and the degree of improvement is observed, and the correction is stopped when the model is fit. Based on completing the assessment of model fit and correction, it is necessary to test the hypothetical paths of the model proposed in this study, construct a model of the influencing factors of learning adaptability, and clarify the relationship between the influencing effects of each factor.

The difference in the lexicalization mode of sports events affects people's choice and attention to the semantic features of sports events and provides certain supporting evidence for the weak hypothesis of linguistic relativity, compared with the nongestural learning group and the multimodal with iconic gesture interpretation. The effect of attitude learning on vocabulary acquisition of students is more significant, indicating that gestures are beneficial to promote learners' cognitive processing and enhance memory representation.

As shown in Figure 5, there are two significant impact paths on multimedia-based interactive learning self-efficacy. First, learning self-efficacy has a direct effect on learning adaptability. This indicates that the more confident students are in their abilities, the better their learning adaptability will be, and thus, they will be able to overcome the difficulties they

encounter in the learning process and achieve good English-learning results. In addition, learning self-efficacy has a significant influence on English-learning motivation and through learning motivation on learning adaptability. In other words, the more self-efficacy students have in the learning process, the stronger their motivation will be, which in turn will increase their learning adaptability and improve their learning outcomes (see Figure 5).

*3.3. Validation of the Model of Factors Influencing the Adaptability of Interactive English Teaching Based on Multimedia Environment.* This study applies a multimedia-based interactive English-teaching adaptation intervention model for college students to English-teaching practice and adopts an action research approach to verify the practical effects of the model. Three specific rounds of action experiments are conducted, and the intervention model and its application are reflected and improved according to the implementation effects of each round, which eventually form a scientific multimedia-based interactive English-teaching adaptation intervention model. This study applies the intervention strategies for learning adaptations to specific multimedia-based interactive English-teaching activities to test the effectiveness of the intervention. Three rounds of action research will be conducted in this study, and each round will include four components: planning, action, observation, and reflection. Through continuous practice, reflection, and revision, the learning adaptability of college students will be gradually improved. The specific design idea of the action research in this study is shown in Figure 6.

After data collection, it is also necessary to use the corresponding tools for the scientific processing of the data. First, according to the process data of this study's intervention, the processing tools corresponding to the data then include the classroom teacher-student behavior analysis coding system, the student English classroom learning engagement level analysis coding system, the SOLO classification-based learning thinking level analysis framework, the classroom teacher-

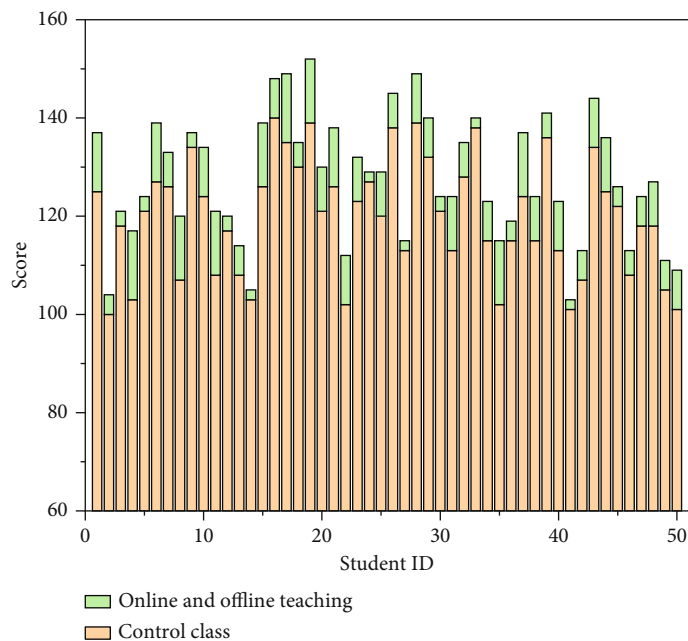


FIGURE 5: Results of the hypothesis testing of the model of factors influencing learning adaptability.

student behavior observation form, the data analysis system in the multimedia backend, and the online discussion content-coding form [28–30]. Among them, the classroom teacher-student behavior observation form was mainly used for the researcher to make written records during the qualitative observation. The other tools were mainly used to record quantitative observations. Second, based on the outcome data of this study's intervention, the processing tools corresponding to the data then included the multimedia-based interactive ELT adaptation scale, the unit knowledge test questions, the reflection level coding form, the student interview outline, and the teacher interview outline.

College students' learning ability and its various dimensions (knowledge acquisition and application ability, learning process self-monitoring ability, learning resource management and application ability) and various dimensions of learning motivation (learning interest, internal motivation, external motivation, achievement motivation), learning self-efficacy. There is a significant correlation. The learning interest dimension and learning self-efficacy in learning motivation have a relatively large impact on the learning ability of college students.

Based on the quantitative analysis of the discussions, a more in-depth analysis of the content of the students' discussions was conducted. Further categorization was done based on the online discussion content-coding sheet designed for this study. By counting, in the first round of intervention, there were 106 items of no substance, 387 items of resource sharing and platform operation, 85 items of insight exchange, 191 items of asking questions, 881 items of answering questions, 58 items of expressing opinions, and 48 items of feedback and comments. As shown in Figure 7, the percentage of the total number of items discussed in each category was further calculated.

#### 4. Multidimensional Analysis of the Path Form of English Teaching in Interactive English Classroom Based on Multimedia Environment

4.1. *Multidimensional Data Analysis of Multimedia Interactive English-Teaching Offline Process.* The offline process data consisted of three sections: teacher-student teaching and learning behaviors, student response level, and student learning engagement level in classroom English teaching. This part mainly uses the classroom teacher-student behavior analysis coding system to analyze a series of behaviors of students and teachers during the class. Specifically, two main aspects of classroom teacher-student behavior distribution and classroom interaction behavior were analyzed. It should be noted here that in the first round of experiments, there were four classes, each of which lasted about 45 minutes. In this study, the classroom recordings were sampled at 3-second intervals, and according to the classroom ELT behavior analysis coding system, there was an average of 900 codes per lesson. In this study, we analyzed each lesson in detail, calculated the percentage of different behaviors, and combined it with the observation of classroom teaching behaviors to determine the structure of ELT and draw conclusions. According to the needs of the study, this section calculates eight aspects (percentages) of teacher language (T), student language (S), silence (C), technology (M), teacher questions (Tq), student responses to questions (Sc), student discussion (Sd), and student questions (Sq). The statistical results of the four lessons in the first round of intervention are shown in Figure 8.

The student speech occupies a certain proportion in all four lessons, which indicates that teachers give students appropriate learning autonomy during lessons. In a comprehensive analysis, the proportion of students' speech can be

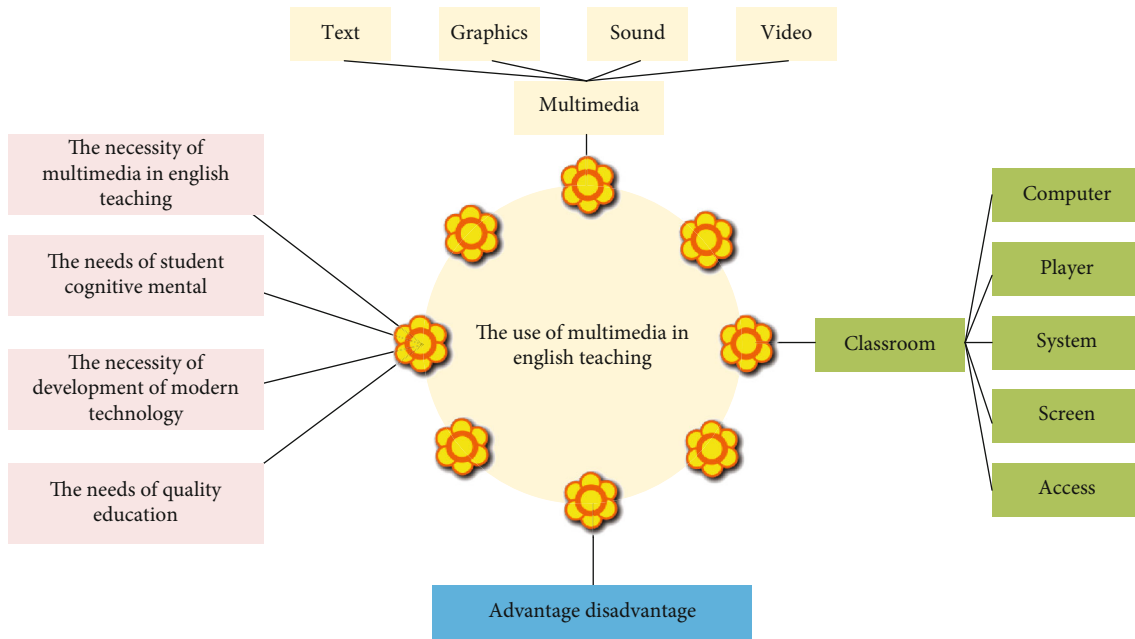


FIGURE 6: Action research design ideas for teaching English in interactive English classrooms.

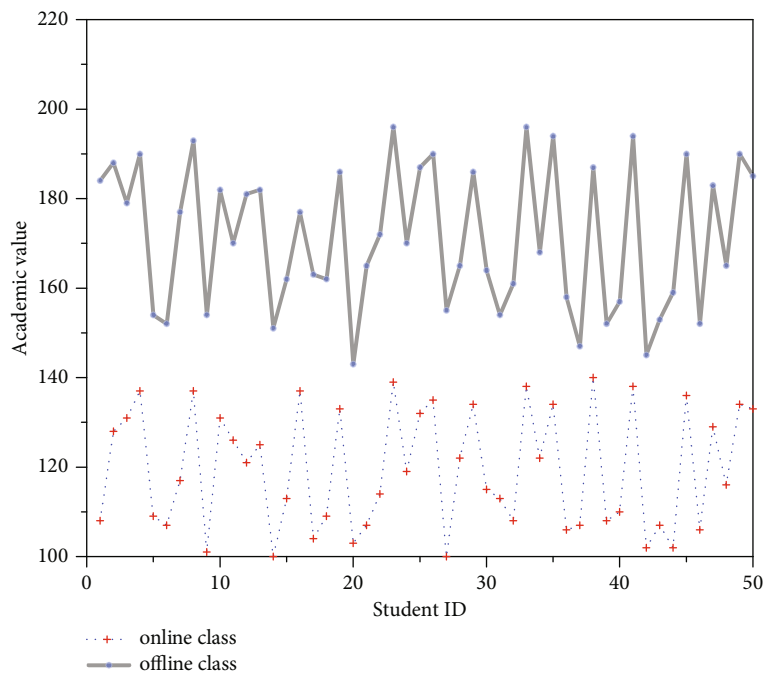


FIGURE 7: Percentage of online discussion content in English classroom English teaching in the multimedia environment.

further increased in subsequent English teaching to guarantee students' learning autonomy. In all four lessons, the teacher's questions will occupy a certain proportion, some thinking questions will be asked for students, and students will actively answer the teacher's questions. In addition, combined with the observation and review of the multimedia classroom videos, two problems can be found. We think the so-called self-learning, to put it in a more popular way, is to learn on your own. Many students do not know how to

learn independently and can only follow the teacher. It is very important for students to learn to learn independently. What we are advocating is lifelong education, that is, to live and learn. This kind of education emphasizes how important learning autonomy is. First, although teachers focus on asking questions to students, they sometimes do not ask students to answer the questions and do not give them time to think. Second, after students answered the questions, they did not provide more in-depth guidance to stimulate students

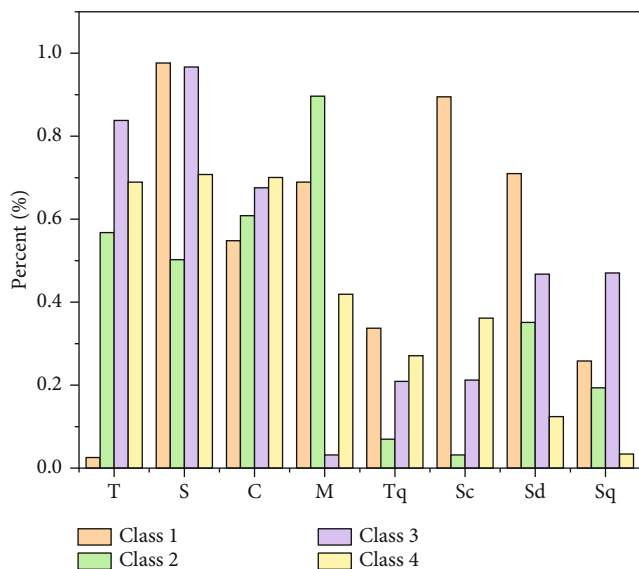


FIGURE 8: Distribution of teacher-student behavior and percentage of classroom interaction behavior in interactive English-teaching English classroom in the multimedia environment.

to think further about the questions. These issues need to be improved later. In contrast, the percentage of students asking questions was significantly small. Students still mainly participated passively and lacked the awareness and spirit of active inquiry and problem identification. In addition, the analysis of the video revealed that the students’ motivation to participate in the classroom needs to be improved; in particular, the motivation to actively answer questions is not high enough, and sometimes, they go to answer questions and discuss at the request of the teacher.

The level of students’ responses to questions is a key measure of student learning. This study analyzes the level of students’ thinking in answering questions in class based on the SOLO thinking level classification framework to determine the level of students’ understanding. Specifically, the number of different response levels in each lesson in the first round was calculated. As shown in Figure 9, the levels of students’ responses to the questions were mainly concentrated on the two levels of single-point structure and multipoint structure (medium level), indicating that most students were able to answer the questions asked by the instructor using basic knowledge and information. The number of higher levels is low in terms of the levels of associative structure and abstract expansion structure (see Figure 9).

The overall aspect was studied for all students, and the overall student learning engagement was observed to calculate the percentage and change in the number of students at each of the three learning engagement levels over time. On this basis, the effects of the intervention were compared across different rounds. Second, on the case study side, the specific learning engagement of a subset of students was analyzed in detail, the changes in learning engagement of these students were followed up continuously, and the effects of the interventions were compared across different rounds. The four sessions were named Video 01, Video 02, Video

03, and Video 04, and each session was 45 minutes in length. To facilitate visual analysis, the study combined the time in one-minute increments (see Figure 10).

As shown in Figure 10, the learning engagement levels of most students remained at the moderate and high engagement levels throughout the ELT process, compared to the relatively small number of students at the nonengaged level, and the change was relatively flat. However, nonengaged students still occupy a certain proportion, and there are relatively more nonengaged students in some periods, which indicates that they appear to have a certain degree of low learning engagement and learning discomfort in the process of learning. It should be noted that the periods with high numbers of high engagement and nonengagement are where the most attention should be paid, so these two parts of the classroom were analyzed primarily. Further analysis revealed that students’ high engagement time was concentrated in minutes 1-4, 11-15, 22-24, 35-36, and 40-42. Combined with the specific situation of the class, in minutes 1-4, the teacher summarizes the online learning and introduces examples and introduces the lesson by asking questions, and students pay more attention. Between 11 and 15 minutes, the teacher conducts special class questions and encourages students to speak actively, and students participate in group discussions and answer questions. At 22-24 minutes, the teacher continues to ask questions and encourages students to take the initiative to ask questions. The teacher emphasizes important points to students and draws their attention to them. At 40-41 minutes, the teacher asks further questions to stimulate students to discuss and think, answer questions, and participate in interactions.

On the other hand, students’ nonengagement time was concentrated in minutes 16-18, 21, 27-29, 31-34, and 44-45. In the context of the lesson, there was an intermittent decrease in students’ attention in minutes 16-18 and 21, after the teacher commented on students’ responses and gave answers. At minutes 27-29 and 31-34, the teacher was mainly lecturing and the content was mainly theoretical, memorized, and conceptual points. At minutes 44-45, when students were answering questions, some students did not go listen attentively. From the above, a certain proportion of low-involvement students appeared in the English teaching classroom, and the duration of high-involvement learning (high number of high-involvement) was not long, and the distribution of high-involvement learning behaviors was discrete and not continuous and concentrated enough.

4.2. *Multidimensional Analysis of Multimedia Interactive English-Teaching Outcome Data.* The outcome data consisted of the learning adjustment level test, unit knowledge test scores, reflection levels, student interviews, and teacher interviews. This section utilizes the learning adaptation scale developed by the study to test for changes in students’ levels of learning adaptability after the intervention. Specifically, a paired-sample *t*-test was used to compare the predicted scores of the experimental students.

To analyze the learning inputs of the 30 students in detail, the study summarized the percentages of different learning input behaviors of these students to the total frequency. The



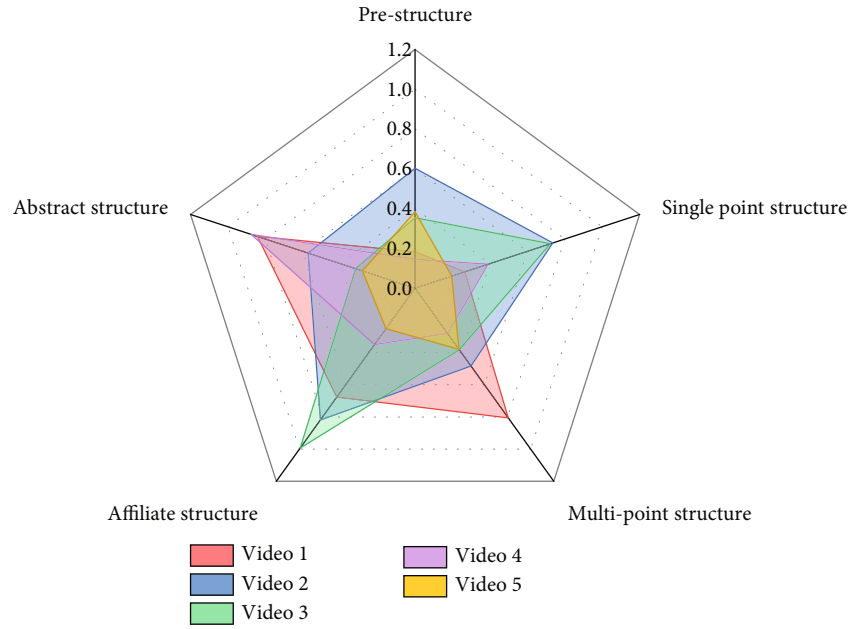


FIGURE 9: Percentage of students' response level in English classroom of interactive English-teaching mode.

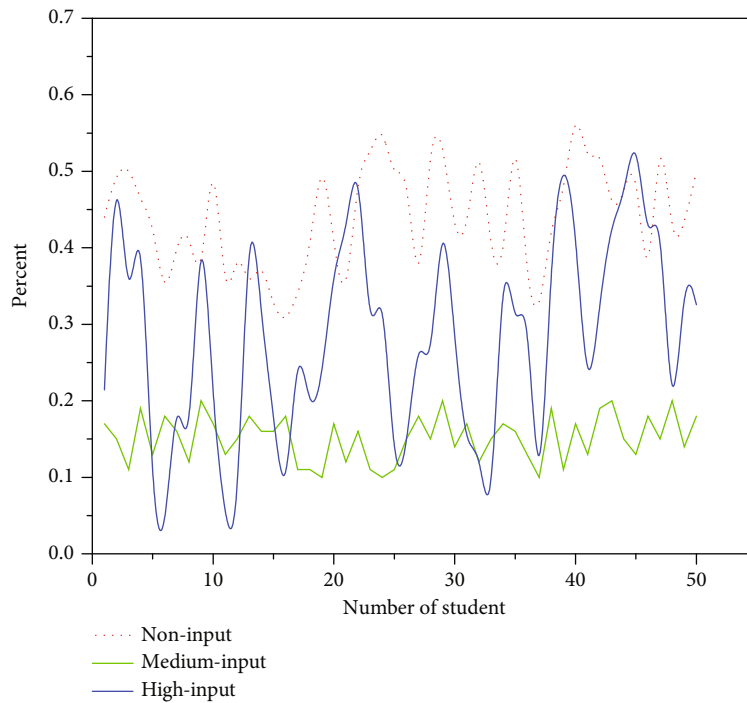


FIGURE 10: Change in the number of people with learning engagement levels over time for multimedia intervention video ELT.

distribution of learning inputs of these students can be seen more visually. Overall, in the first round of the experiment, most students had a larger percentage of moderate input behaviors versus high input behaviors. The largest proportion of moderate input behaviors indicates that moderate input was the most dominant learning input behavior among the 30 students. After that, it was the high-involvement behavior. On further analysis, the graph reveals the problem that the percentage of high-involvement behaviors of students is relatively

low. The percentage of high-involvement behaviors of these students is only 18.14%, 13.58%, 19.69%, 21.86%, 21.64%, and 17.28%, respectively, which should be focused on. On the other hand, all 30 students showed noninvolving learning behaviors. Among them, several students had a high percentage of noninvolving behaviors. In particular, these students had 17.39%, 15.80%, 22.64%, 25.69%, 20.61%, 16.17%, and 18.14% of noninvolving behaviors. Some students showed a significantly low percentage of high-involvement behaviors or

a high percentage of noninvolvement behaviors, which needs to be improved in the next round of the experiment.

To more visually reflect the change in the level of reflection of the students in different rounds, Figure 1 shows the percentage of the number of students who were at different levels with no intervention and the first round and the second round of intervention. According to the graph, the number of students at low levels of reflection increased slightly through the second round of intervention. The number of people at medium to high levels of reflection increased further. This requires further strengthening of reflection instruction for students in the next round of intervention to reduce the number of low-level reflections. This study conducted three rounds of action research to verify the effectiveness of the intervention model by designing appropriate learning resources and learning activities and conducting English-teaching practices based on the learning adaptation intervention model. In each round of the study, the four steps of action research (planning, action, observation, and reflection) were strictly followed. Through the collecting process and the outcome data and corresponding data analysis, reflection was conducted on this basis to summarize the effects achieved and the shortcomings. Based on the results of reflection, the learning adaptation intervention model is further revised to provide the basis for the next round of research. This cycle of three rounds of action research basically achieved the goals set by the study, solved the corresponding practical problems, gradually improved the learning adaptability of college students, and optimized the learning effect.

## 5. Conclusion

This paper further defines the composition by developing a scale of English language-teaching adaptability and its influencing factors in English classrooms based on a multimedia-based interactive English language-teaching model through the methods of item analysis, exploratory factor analysis, and validating factor analysis. First, the learning adaptability scale was developed, which can be applied to measure the level of adaptation to multimedia-based interactive English language teaching among college students, and further defines the composition of multimedia-based interactive English language-teaching adaptability. Secondly, the development of the learning adaptation influencing factor scale can be used to further define the composition of the influencing factors of multimedia-based interactive English-teaching adaptation and lay the foundation for exploring the more complex role relationships among the influencing factors of English classroom English-teaching-learning adaptation and improving college students' learning adaptability.

This study strictly follows the four steps of action research and further revises the learning adaptation intervention model based on the reflection results, to gradually improve college students' adaptation to multimedia-based interactive English-teaching English classrooms and optimize the learning effect. First, the data is processed online, the number of student discussions gradually increased, and the proportion of discussion content such as active questions and expressing opinions gradually increased. On the offline

side, the percentage of students' speech gradually increased, the quantity and quality of students' answers to questions improved, and the number of students' active questions in English classroom English teaching also gradually increased and their motivation continued to improve. The situation that high learning engagement behaviors are not long-lasting, unfocused, and discrete is gradually improving. The percentage of students' high engagement behaviors gradually increased, and the percentage of nonengagement behaviors gradually decreased. Second, the outcome data, student's level of learning adaptability, unit knowledge mastery, and reflection gradually improved over the multiple rounds of intervention.

This study developed a multimedia-based evaluation of English language-teaching adaptations in interactive English language-teaching English classrooms, which was partly reflected in the validation of the effectiveness of the intervention model of English language-teaching-learning adaptations in English classrooms. The students' process data were collected and analyzed, including the number and content of online discussions, offline classroom teachers' and students' teaching and learning behaviors, offline students' question answering levels, and offline classroom students' learning engagement levels. We collected and analyzed the outcome data of students' learning in English classrooms, including the learning adaptation level test, the unit knowledge level test, the reflection level test, and the teacher-student interviews. By comprehensively collecting students' process data and outcome data, we can not only give students scientific evaluation but also identify problems and correct intervention strategies in time, which can provide some reference for future research related to interactive English classroom English teaching in a multimedia environment.

## Data Availability

All information is within the paper.

## Conflicts of Interest

No competing interests exist concerning this study.

## References

- [1] L. Zhao, L. Chen, Q. Liu, M. Zhang, and H. Copland, "Artificial intelligence-based platform for online teaching management systems," *Journal of Intelligent & Fuzzy Systems*, vol. 37, no. 1, pp. 45–51, 2019.
- [2] N. Zhao, X. Zhou, B. Liu, and W. Liu, "Guiding teaching strategies with the education platform during the COVID-19 epidemic: taking Guiyang no. 1 middle school teaching practice as an example," *Science Insights Education Frontiers*, vol. 5, no. 2, pp. 531–539, 2020.
- [3] A. Alam, "Possibilities and challenges of compounding artificial intelligence in India's educational landscape," *International Journal of Advanced Science and Technology*, vol. 29, no. 5, pp. 5077–5094, 2020.
- [4] C. Shen, "MOOC teaching mode of news transmission based on network audio data decoding technology," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 13, no. 6, pp. 43–55, 2018.

- [5] S. Li, "Innovations in Chinese engineering education with digital technologies: a brief review of recent advances," *Computer Applications in Engineering Education*, vol. 26, no. 5, pp. 1081–1088, 2018.
- [6] W. Liu, H. L. Ma, and A. Walsh, "Advance in photonic crystal solar cells," *Renewable and Sustainable Energy Reviews*, vol. 116, p. 109436, 2019.
- [7] X. Zhang, C. Zang, H. L. Ma, and Z. J. Wang, "Study on removing calcium carbonate plug from near wellbore by high-power ultrasonic treatment," *Ultrasonics Sonochemistry*, vol. 62, p. 104515, 2020.
- [8] H. L. Ma, X. Zhang, F. F. Ju, and S. B. Tsai, "A study on curing kinetics of nano-phase modified epoxy resin," *Scientific Reports*, vol. 8, no. 1, p. 3045, 2018.
- [9] F. Qiao and H. Wang, "Mobile interactive translation teaching model based on "Internet+"," *EURASIA Journal of Mathematics, Science and Technology Education*, vol. 13, no. 10, pp. 6705–6714, 2017.
- [10] L. H. Li, "Design of college English process evaluation system based on data mining technology and internet of things [J]," *International Journal of Data Warehousing and Mining (IJDWM)*, vol. 16, no. 2, pp. 18–33, 2020.
- [11] D. Zou, H. Xie, and F. L. Wang, "Future trends and research issues of technology-enhanced language learning: a technological perspective [J]," *Knowledge Management & E-Learning: An International Journal*, vol. 10, no. 4, pp. 426–440, 2018.
- [12] M. Ling, M. J. Esfahani, H. Akbari, and A. Foroughi, "Effects of residence time and heating rate on gasification of petroleum residue," *Petroleum Science and Technology*, vol. 34, no. 22, pp. 1837–1840, 2016.
- [13] H. L. Ma and S. B. Tsai, "Design of research on performance of a new iridium coordination compound for the detection of Hg<sup>2+</sup>," *International Journal of Environmental Research and Public Health*, vol. 14, no. 10, p. 1232, 2017.
- [14] L. Y. Mo, W. H. Z. Sun, S. Jiang et al., "Removal of colloidal precipitation plugging with high-power ultrasound," *Ultrasonics Sonochemistry*, vol. 69, p. 105259, 2020.
- [15] D. Xu and H. Ma, "Degradation of rhodamine B in water by ultrasound-assisted TiO<sub>2</sub> photocatalysis," *Journal of Cleaner Production*, vol. 313, p. 127758, 2021.
- [16] M. Ivanova, "eLearning informatics: from automation of educational activities to intelligent solutions building," *Informatics in Education*, vol. 19, no. 2, pp. 257–282, 2020.
- [17] J. Q. Li, F. R. Yu, G. Deng, C. Luo, Z. Ming, and Q. Yan, "Industrial internet: a survey on the enabling technologies, applications, and challenges," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 3, pp. 1504–1526, 2017.
- [18] S. Zou, "Designing and practice of a college English teaching platform based on artificial intelligence," *Journal of Computational and Theoretical Nanoscience*, vol. 14, no. 1, pp. 104–108, 2017.
- [19] D. Gao, Y. Liu, Z. Guo et al., "A study on optimization of CBM water drainage by well-test deconvolution in the early development stage," *Water*, vol. 10, no. 7, p. 929, 2018.
- [20] S. B. Tsai and H. Ma, "A research on preparation and application of the monolithic catalyst with interconnecting pore structure," *Scientific Reports*, vol. 8, no. 1, p. 16605, 2018.
- [21] J. Xie and H. Ma, "Application of improved APO algorithm in vulnerability assessment and reconstruction of microgrid," *Iop Conference*, vol. 108, no. 5, article 052109, 2018.
- [22] S. Dai, "ARS interactive teaching mode for financial accounting course based on smart classroom," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 14, no. 3, pp. 38–50, 2019.
- [23] F. Kong, J. Li, and Y. Wang, "Human-computer interactive teaching model based on fuzzy set and BP neural network," *Journal of Intelligent & Fuzzy Systems*, vol. 37, no. 1, pp. 103–113, 2019.
- [24] J. Kong, "Innovative applications mode of network learning space in exercise physiology based on ubiquitous learning," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 14, no. 4, pp. 113–126, 2019.
- [25] Y. Zhang, Z. Zhou, H. Bai, W. Liu, and L. Wang, "Seizure classification from EEG signals using an online selective transfer TSK fuzzy classifier with joint distribution adaption and manifold regularization," *Frontiers in Neuroscience*, vol. 14, pp. 1–9, 2020.
- [26] Y. Zhang, F. L. Chung, and S. Wang, "Fast exemplar-based clustering by gravity enrichment between data objects," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 50, no. 8, pp. 2996–3009, 2020.
- [27] Y. Zhang, F. L. Chung, and S. Wang, "Takagi-Sugeno-Kang fuzzy systems with dynamic rule weights," *Journal of Intelligent & Fuzzy Systems*, vol. 37, no. 6, pp. 8535–8550, 2019.
- [28] T. Grubljesic, P. S. Coelho, and J. Jaklic, "The shift to socio-organizational drivers of business intelligence and analytics acceptance," *Journal of Organizational and End User Computing*, vol. 31, no. 2, pp. 37–64, 2019.
- [29] L. X. Z. Zhang, M. Mouritsen, and J. R. Miller, "Role of perceived value in acceptance of bring your own device policy," *Journal of Organizational and End User Computing*, vol. 31, no. 2, pp. 65–82, 2019.
- [30] A. Shahri, M. Hosseini, K. Phalp, J. Taylor, and R. Ali, "How to engineer gamification: the consensus, the best practice and the grey areas," *Journal of Organizational and End User Computing*, vol. 31, no. 1, pp. 39–60, 2019.

## Research Article

# Flipped Classroom for Motor Skills: What Factors Influence College Students' Learning Effect?

Fengyan Zhang,<sup>1</sup> Baojuan Ma ,<sup>1</sup> and Wengang Ren<sup>2</sup>

<sup>1</sup>Physical Education Department, Shijiazhuang Information Engineering Vocational College, Shijiazhuang, 050000 Hebei, China

<sup>2</sup>Department of Sports Training, Hebei Sport University, Shijiazhuang, 050000 Hebei, China

Correspondence should be addressed to Baojuan Ma; mabaojuan2021@163.com

Received 13 July 2021; Revised 30 July 2021; Accepted 10 August 2021; Published 19 August 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Fengyan Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The study of motor skills is an essential course item for college students. Many college teachers try to introduce the flipped classroom mode into the teaching of motor skills. However, the influence of this mode on the learning effect of college students' motor skills is uncertain. This paper builds a research model on the basis of literature review. This model is used to analyze the influence factors of flipped classroom on the learning effect of college students' motor skills. Through questionnaire survey and empirical analysis, we verified the research hypothesis of each factor. The results show that video technical action display, video feedback of students' technical action, and teaching interaction have a significant positive impact on the learning effect of college students' motor skills in the flipped classroom, while video theory teaching has no significant impact. This study expands the application of flipped classroom in the field of motor skills and enriches the relevant literature of flipped classroom mode. Based on the findings of this study, we make the following recommendations: (1) teachers can use the way of video recording to display technical actions of motor skills, (2) students can use video feedback to communicate with the teacher about the learning of motor skills, and (3) flipped classroom needs to pay attention to the teaching interaction between teachers and students.

## 1. Introduction

In the process of physical education in colleges and universities, the study of motor skills is becoming more and more important. Motor skills refer to the ability of human body to master and effectively complete special actions in the process of movement [1]. Students' master motor skills, on the one hand, can improve the scientific nature of sports and, on the other hand, can prevent unnecessary injuries [1, 2]. Therefore, paying attention to the training of students' motor skills is an essential goal in physical education teaching. Physical education is a practical course, and the study of theoretical knowledge is equally important. Because mastering complete physical education theoretical knowledge allows students to get scientific guidance while engaging in physical activity [3].

With the progress of information [4–6] and communication technology [7, 8] and the wide application of mobile

smart phones [9], the traditional offline teaching method is gradually changing, and online teaching [10] has been introduced into the training of motor skills courses for college students. Inspired by the thinking of online teaching, a teaching mode called flipped classroom has been applied to college students' motor skills courses. Flipped classroom refers to the realignment of time in and out of the classroom and the transfer of learning decisions from teachers to students [11, 12]. This new form of education and teaching overturns the traditional classroom teaching mode, and students gain the initiative instead of passively receiving education. This is not only a kind of role transformation but also the innovation of education mode [13].

It is conceivable that under the flipped classroom mode, students can independently learn theoretical knowledge of motor skills through video teaching at home, dormitory, or cafe [14]. In addition, students can play the course video repeatedly to get an in-depth understanding of the

characteristics of motor skills and action process and eventually form an internal movement image and correct movement concept in their mind, laying a foundation for formal practice [15]. For some exercise tasks of motor skills after class, students can submit them to their teachers by recording videos [16]. This not only plays a role in testing students' mastery of motor skills but also helps teachers fully understand each student's learning situation. The learning of motor skills is a compulsory course for college students. Many college teachers try to introduce the flipped classroom model into motor skills teaching, and the effect of this model on the learning effect of college students' motor skills is still uncertain. Therefore, this paper constructs a research model to expand the application of flipped classrooms in the field of motor skills.

However, it is controversial whether the flipped classroom mode can enhance students' learning outcomes in general. For example, Persike's [16] research shows that in flipped classrooms, one of the determinants of rising course performance is having high-quality traditional courses. When a student's course performance is already high, there may be little room for further improvement in the flipped classroom mode. Clifton and Boutell [17] put forward that flipped classroom has different learning effects for students with different learning abilities. For example, students who learn quickly and have a strong understanding ability in a flipped classroom will return to a traditional classroom feeling confident and experienced but bored by it. Those students who did not learn well enough in the flipped classroom and did not fully grasp the knowledge of the course often felt struggling and at a loss when they returned to the traditional classroom. Therefore, this study tries to answer what are the factors that affect the learning effect of college students' motor skills in flipped classroom and what are their mechanisms of influence?

Given the above analysis, we find that flipped classrooms, while having many advantages, are not without drawbacks. Therefore, it is of great significance to study the factors that may influence the learning effect of students in the flipped classroom mode. On the one hand, this study is conducive to stimulating students' enthusiasm to participate in flipped classroom, and on the other hand, it is conducive to increasing the interaction between students and teachers in motor skills of flipped classroom. In addition, this study takes motor skills learning as the research object, and explores the potential factors that affect students' learning effect under the flipped classroom mode. This study enriches the literature of flipped classroom and provides theoretical guidance for the application of motor skills in flipped classroom.

The rest of this paper is arranged as follows: the second section is literature review, the third section proposes the model construction and research hypothesis, the fourth section presents the research method and empirical results, and finally, the fifth section puts forward the conclusions.

## 2. Literature Review

*2.1. Motor Skills.* Motor skills are also called "action skills." It refers to the ability of human body to master and effectively

complete special actions in the process of movement [1, 18]. In the teaching of physical education, the theory and practice of motor skills are very important [19]. Traditional physical education courses are generally carried out in an offline way. For example, in the first half of the course, the teacher explains the theoretical knowledge related to motor skills to the students, while in the second half of the course, the students practice motor skills [20]. With the development of information and communication technology and the wide application of mobile smart phones, network learning has been introduced into physical education [21]. In fact, Williams and Grant [22] have pointed out that video technology or online simulation teaching methods may be effective for the development of students' motor skills. Especially when the network technology is combined with the teaching technology, it is beneficial to the development of students' physical awareness.

Generally, the formation of students' motor skills needs to go through three stages [23, 24]: the cognitive stage, the connection stage, and the perfection stage. The first is the cognitive stage. At this stage, the practitioner mainly demonstrates the action through visual observation and mimics the exercise, using vision to control the action [25]. After a certain amount of practice, the practitioner has initially mastered a series of local actions and began to connect the independent movements. At this point, the practitioner enters the connection phase. In this stage, the practitioner's attention mainly points to the details of skills, generalizes the essential characteristics of movements through thinking analysis, gradually realizes the whole movement, and combines several individual actions into a whole [26]. Finally, there is the perfection stage. At this stage, the movements of the practitioner have established a consolidated dynamic pattern in the brain, a series of movements mastered have formed a complete organic system, and each action can be expressed in the form of linkage [27].

*2.2. The Learning Method of Motor Skills in Flipped Classroom.* The introduction of flipped classroom innovates the traditional physical education teaching method. Especially when it comes to motor skills, students have more autonomy and can arrange their time freely [11]. For teachers, they can record the theoretical knowledge and action points about motor skills in advance [12]. They can also perform physical demonstrations of a set of motor skills, which can also be recorded on video. For students, they can learn theoretical knowledge and action points through videos. Actions can also be imitated through the video provided by the teacher [13]. In fact, testing students' practicality is very effective in flipped classroom. For example, after every lesson, teachers can ask students to record videos of their own actions. The videos are then sent to the teacher. Teachers can judge students' mastery of motor skills through the videos provided by students [14]. It can be seen that the application of flipped classroom in physical education can, on the one hand, impart knowledge and motor skills and, on the other hand, effectively supervise and familiarize students with their learning effects in motor skills. However, some scholars have pointed out that such flipped video

teaching may be detrimental to the interaction between teachers and students [11].

Through literature review, we found that scholars have provided us with a rich literature basis for motor skills and flipped classroom. As a product of the development of modern science and technology, flipped classroom is becoming more and more popular in college teaching. Especially in college physical education courses, flipped classroom can realize the combination of theory and practice teaching. However, there is still debate about the impact of flipped classroom on student learning outcomes (e.g., Persike [16]; Clifton and Boutell [17]). Therefore, in the field of motor skills, we propose it is necessary to further explore the effect of this teaching mode. The study on this topic, on the one hand, can provide theoretical reference for the application of flipped classroom in physical education teaching and, on the other hand, can provide improvement recommendations for students in the learning of motor skills.

### 3. Model Construction and Hypothesis

*3.1. Model Construction.* We construct a model of the factors influencing the learning effect of college students' motor skills based on a literature review of flipped classrooms. As depicted in Figure 1, the four components of this theoretical model are video theory teaching, video technical action display, video feedback of students' technical actions, and teaching interaction.

#### 3.2. Hypothesis

*3.2.1. Video Theory Teaching.* The teaching of theoretical knowledge is an important link of physical education curriculum. Usually, in the first few classes of the whole course, or in the first few minutes of each class, the teacher will teach the students the theoretical knowledge of motor skills. In the flipped classroom, the theoretical knowledge can be recorded into a video in advance. Students can study independently according to the schedule of the course [11]. It is worth noting that these videos can be watched over and over again. This helps students to strengthen their understanding of the knowledge points. And even if the students do not understand the point, they can also ask the teacher in the offline teaching. Therefore, video teaching is helpful for students to deepen their understanding of theoretical knowledge of motor skills in advance. But other scholars have raised concerns. For example, if students' autonomous learning ability is not strong, video learning is difficult to ensure students' learning effect. In addition, students learn theoretical knowledge in advance through videos and may feel bored in offline courses [17]. Therefore, it is very important to understand the influence of video theory teaching on students' learning effect. We choose one of these possibilities to hypothesize:

H1: There is a positive correlation between the video theory teaching in flipped classroom and the learning effect of college students' motor skills.

*3.2.2. Video Technical Action Display.* Flipped classroom can realize the video of various teaching contents. In physical

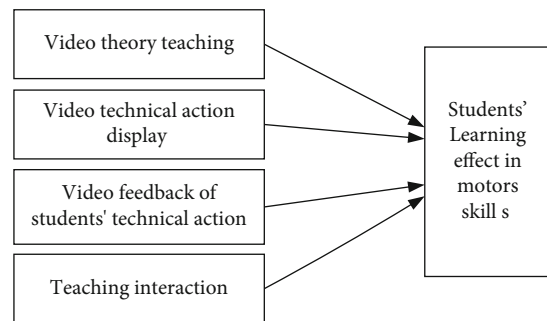


FIGURE 1: Research theoretical model.

education teaching, the imparting of motor skills is very important. It not only needs to teach theoretical knowledge to students but also needs to carry on the technical action demonstration. In offline classes, teachers teach students by demonstrating technical movements over and over again. However, this approach poses a great challenge to teachers' teaching. The application of video technology in flipped classroom reduces the teaching burden of teachers [12]. Teachers can record standard technical actions on video, and students can watch the videos over and over again. Especially for some difficult actions, students can slow down the playing speed of the video or pause the video for a specific action [14]. This is helpful for students to decompose the technical actions and then quickly understand and learn the technical actions. Based on the above discussion, we propose the following hypothesis:

H2: There is a positive correlation between the display of video technical actions in flipped classroom and the learning effect of college students' motor skills.

*3.2.3. Video Feedback of Students' Technical Action.* One way to test students' mastery of motor skills is to have them record videos. After teachers teach theoretical knowledge and technical actions, students need to master these points through repeated practice. However, in order to supervise the learning effect of students and to understand the learning implementation of each student, teachers can ask students to record videos for feedback [16]. This approach has the following advantages [13]: one is to give students the autonomy to submit a perfect work when they think they have practiced it perfectly; second, teachers play the role of course supervision and can understand the students' learning situation. Accordingly, we put forward the following research hypothesis:

H3: There is a positive correlation between the video feedback of students' technical actions and the learning effect of college students' motor skills.

*3.2.4. Teaching Interaction.* Teaching interaction has always been regarded as an important part of curriculum learning. The advantage of it is that it can activate the classroom atmosphere, enhance students' learning attention, deepen students' understanding of knowledge points, and help teachers to understand students' learning situation. However, flipped classroom may have a few disadvantages in

TABLE 1: Scale of observed variables.

Latent variables	Observed variables	References
Video theory teaching (VTT)	1a: I like to learn the theoretical knowledge of motor skills through videos. 1b: The teaching method of video theory has the advantage of repeated play. 1c: The teaching method of video theory enables me to arrange my learning time independently.	Herreid and Schiller [11]; Clifton and Boutell [17]
Video technical action display (VTAD)	2a: I like the way of video technical action display. 2b: This way of learning allows me to deepen my understanding of each technical action. 2c: This learning method provides me with a good opportunity to imitate practice.	Herreid et al. [12]; Mason et al. [14]
Video feedback of students' technical action (VFSTA)	3a: I think this way of learning supervision has a strong autonomy. 3b: I prefer to practice the technique very well and then record the video and send it to the teacher. 3c: The existence of such feedback mode makes me constantly remind myself to carefully study the theoretical knowledge and technical actions in the course.	Pierce and Fox [13]; Persike [16]
Teaching interaction (TI)	4a: In the process of video learning, I can only solve any question by myself. 4b: I wish there was a better way to help me answer my questions during the video learning process. 4c: Due to the lack of real-time interaction with the teacher, I felt the process of independent learning was boring.	Herreid and Schiller [11]; Cheng and Wang [15]
Students' learning effect in motors skills (SLEMS)	5a: I think the video teaching in flipped classroom will affect the effect of students in the process of learning motor skills. 5b: I think it is necessary for students to give video feedback in the process of learning motor skills. 5c: I think flipped classroom also needs to pay attention to the interaction between teachers and students.	Persike [16]; Clifton and Boutell [17]

TABLE 2: Descriptive statistical analysis of the survey results.

Basic information	Items	Rate	Basic information	Items	Rate
Age	15-20	8%	Whether take courses in motor skills	Yes	100%
	20-25	90%		No	0%
	>25	2%			
Gender	Male	89%	Whether know about flipped classroom	Yes	100%
	Female	11%		No	0%

terms of teaching interaction [11]. For example, if students have questions when learning recorded videos, it is difficult for them to solve them by themselves [15]. If students take notes on these questions and wait until the offline class to ask the teacher for answers, new problems may arise. For example, it is not good for students to grasp knowledge points well in the video learning stage, which may reduce their learning enthusiasm. In addition, if a large number of students have questions and the number of questions generated is also large, this may bring great challenges to the teacher's offline courses. Therefore, we propose the following hypothesis:

H4: If the teaching interaction in flipped classroom can be increased, it will promote the learning effect of college students' motor skills.

## 4. Methodology and Results

*4.1. Questionnaire Design.* We used a questionnaire design to first understand the impact of flipped classroom teaching mode on the learning effect of college students' motor skills. The questionnaire is divided into two sections. The first section contains basic information about the respondents, such as their age, gender, whether they are aware of the flipped classroom, and whether they are enrolled in motor skills courses. For this study, the basic information was used as a control variable. Video theory teaching, video technical action display, video feedback of students' technical action, teaching interaction, and students' learning effect in motor skills are the five latent variables in the second part. We used

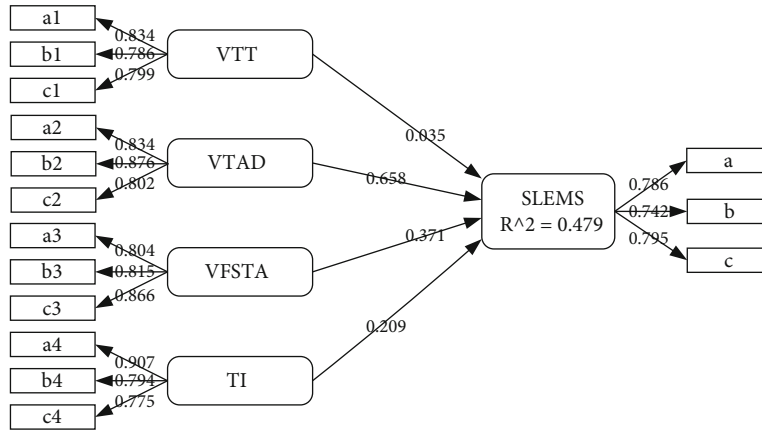


FIGURE 2: Model results.

a 5-point Likert scale to quantify the observed variables of each latent variable [28, 29]. The scale of the second part of the questionnaire is shown in Table 1.

**4.2. Data Collection.** We conducted a pretest on the designed questionnaire. The specific operation process is as follows: first, we invited three students from the motor skills course to fill in a questionnaire; second, according to their answers and the brief interview (about 5 minutes for each student), we found out the possible mistakes in the questionnaire; finally, the questionnaire was revised according to the feedback of the three testers. Our questionnaires were distributed online, and 96 questionnaires were finally recovered, 87 of which were valid, and the effective rate of questionnaire recovery was 90.6%. The survey was conducted over three weeks from May 2021. All respondents were told before taking part that the survey was anonymous and that the data would only be used for scientific research. The respondents were all from our universities, and they all took the course of motor skills. Moreover, this is a public welfare investigation. Respondents are willing to participate, and we do not offer any rewards. Descriptive statistical results of the questionnaire are shown in Table 2.

**4.3. Data Analysis.** We used SmartPLS software to conduct empirical analysis on the data [30]. Its principle is to use the method of PLS-SEM to build the influence factor model of flipped classroom on the learning effect of college students' motor skills. PLS-SEM model has good advantages in the treatment of small samples and the verification of the relationship between factors. In recent years, this method has been widely used in the study of influencing factors. For example, Ali and Omar [31] applied PLS-SEM model to evaluate the influence of physical environment and social environment on customer experience. Huang et al. [32] used PLS-SEM model to study the influencing factors of the crowd workers' continuous participation intention in crowdsourcing logistics. Buitrago et al. [33] used PLS-SEM model to explore the relationship between institutional quality and international competitiveness.

TABLE 3: Results of reliability and validity tests.

Construct	CA	CR	AVE
VTT	0.715	0.756	0.648
VTAD	0.923	0.975	0.814
VFSTA	0.821	0.883	0.756
TI	0.709	0.743	0.637
SLEMS	0.820	0.874	0.795

VTT: video theory teaching; VTAD: video technical action display; VFSTA: video feedback of students' technical action; TI: teaching interaction; SLEMS: students' learning effect in motors skills.

TABLE 4: Average variance extracted (AVE) square root and factor correlation coefficient.

	H	H1	H2	H3	H4
SLEMS	<b>0.956</b>				
VTT	0.379	<b>0.803</b>			
VTAD	0.458	0.225	<b>0.752</b>		
VFSTA	0.602	0.197	0.492	<b>0.749</b>	
TI	0.473	0.248	0.660	0.371	<b>0.680</b>

Diagonal elements are the square root of average variance extracted (AVE).

**4.3.1. Reliability and Validity Tests.** We used SmartPLS software to verify the reliability and validity of the five latent variables. The results show that the values of Cronbach's alpha (CA), combined reliability (CR), and average variance extracted (AVE) all reach corresponding thresholds [34, 35]. Meanwhile, the values of external loads for the observation variables are all greater than 0.7 (see Figure 2). Therefore, it can be considered that the model and scale constructed in this study have high reliability and validity [34, 35]. In addition, the value of  $R^2$  is 0.479, indicating that the model we constructed has good explanatory power [35]. According to the data analysis results in Table 3 and Table 4, we find that the square root of AVE is greater than its correlation coefficient with other latent variables. This indicates that there is no multicollinearity among the latent variables [35]. Moreover, the test of model fit shows that the SRMR



TABLE 5: Results of hypothesis testing.

Hypothesis	Path	Path coefficient	<i>t</i> value	<i>P</i> value	Hypothesis supported?
H1	VTT → SLEMS	0.035	0.463	0.735	N
H2	VTAD → SLEMS	0.658	8.342	0.000	Y
H3	VFSTA → SLEMS	0.371	2.985	0.003	Y
H4	TI → SLEMS	0.209	2.679	0.012	Y

value is 0.042, less than 0.08. Therefore, the model constructed in this study has good fitting validity [33, 35].

**4.3.2. Hypothesis Testing.** The bootstrapping method of SmartPLS software was used to verify the research hypothesis. The results of the testing are shown in Table 5. Therefore, hypotheses H2, H3, and H4 passed hypothesis verification ( $P \leq 0.05$ ), while H1 did not ( $P > 0.5$ ) [34, 35]. In other words, video technical action display, video feedback of students' technical action, and teaching interaction have a significant positive impact on the learning effect of college students' motor skills in the flipped classroom, while video theory teaching has no significant impact. Our findings (H2, H3, and H4) have supported the viewpoints and conclusions from Herreid and Schiller [11], Herreid et al. [12], Pierce and Fox [13], Mason et al. [14], Cheng and Wang [15], and Persike [16]. However, the finding (H1) has not confirmed the positive or negative correlation between video theory teaching and the learning effect of students' motor skills in flipped classroom. Therefore, on the hypothesis of video theory teaching, our finding is inconsistent with Herreid and Schiller [11] and Clifton and Bou-tell [17].

## 5. Conclusions

Once the flipped classroom mode was proposed, it attracted the attention of many universities and teaching staff, and more and more courses were taught in this way. The study of motor skills is an essential project for college students. Many college teachers also try to introduce the flipped classroom mode into the teaching of motor skills. However, the influence of this mode on the learning effect of college students' motor skills is uncertain. Accordingly, we built a research model on the basis of literature review. This model is used to analyze the influence factors of flipped classroom mode on the learning effect of college students' sports skills. We focus on four factors, which are video theory teaching, video technical action display, video feedback of students' technical action, and teaching interaction. Through questionnaire survey and empirical analysis, we verify the research hypothesis of each factor. The results show that video technical action display, video feedback of students' technical action, and teaching interaction have a significant positive impact on the learning effect of college students' motor skills in the flipped classroom, while video theory teaching has no significant impact.

This study expands the application of flipped classroom in the field of motor skills and enriches the relevant literature of flipped classroom mode. Based on the findings of this

study, we make the following recommendations: (1) teachers can use the way of video recording to display technical actions of motor skills, (2) students can use video feedback to communicate with the teacher about the learning of motor skills, and (3) flipped classroom needs to pay attention to the teaching interaction between teachers and students.

However, there are some limitations in this paper. First, we used a literature review to validate four potential influencing factors. We may need to consider additional factors depending on the viewpoint: students' learning autonomy and teachers' preferred teaching methods, for example. As a result, future studies can conduct factor analysis and validation from a variety of perspectives. Second, the questionnaire data in this paper comes from motor skills course student groups, and the sample size is small. Large sample studies should be conducted in the future, if conditions allow, to increase the universality of research findings. Finally, a motor skills class may differ from a typical social science or science class. As a result, the flipped classroom model's application effect in various courses may differ. As a result, interdisciplinary comparative research will be necessary in the future.

## Data Availability

The data used to support the findings of this study are included within the article.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

- [1] J. R. Thomas, K. E. French, and C. A. Humphries, "Knowledge development and sport skill performance: directions for motor behavior research," *Journal of Sport Psychology*, vol. 8, no. 4, pp. 259–272, 1986.
- [2] S. Pancar, "The effect of open and closed-skill sports on cognitive functions," *Spor ve Eğitim Bilimleri Dergisi*, vol. 7, no. 7-2, pp. 159–166, 2020.
- [3] A. Yz, C. Ctc, and D. Ngma, "The effects of visual training on sports skill in volleyball players-Science Direct," *Progress in Brain Research*, vol. 253, pp. 201–227, 2020.
- [4] W. Cai, Z. Wei, R. Liu, Y. Zhuang, Y. Wang, and X. Ning, "Remote sensing image recognition based on multi-attention residual fusion networks," *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 1, pp. 1–8, 2021.

- [5] M. Zhao, C. H. Chang, W. Xie, Z. Xie, and J. Hu, "Cloud shape classification system based on multi-channel CNN and improved FDM," *IEEE Access*, vol. 8, pp. 44111–44124, 2020.
- [6] J. Zhang, Y. Liu, H. Liu, and J. Wang, "Learning local-global multiple correlation filters for robust visual tracking with Kalman filter redetection," *Sensors*, vol. 21, no. 4, p. 1129, 2021.
- [7] L. Zhang, X. Wang, X. Dong, L. Sun, W. Cai, and X. Ning, "Finger vein image enhancement based on guided tri-Gaussian filters," *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 1, pp. 17–23, 2021.
- [8] Z. Zhang, C. Fan, L. Zhang, and M. Kong, "K algorithm for microstrip peocessor design," *ASP Transactions on Neural Information Computing*, vol. 1, no. 1, pp. 18–21, 2021.
- [9] H. Y. Choi, T. H. Kim, and J. H. Lee, "Relational analysis of smartphone sports application usage motivation, sports attitude, leisure satisfaction, exercise flow, sport continuance," *Korean Journal of Sports Science*, vol. 26, no. 3, pp. 551–564, 2017.
- [10] L. Liang, Q. Yin, and C. Shi, "Exploring proper names online and its application in English teaching in university," *ASP Transactions on Computers*, vol. 1, no. 1, pp. 24–29, 2021.
- [11] C. F. Herreid and N. A. Schiller, "Case studies and the flipped classroom," *Journal of College Science Teaching*, vol. 42, no. 5, pp. 62–66, 2013.
- [12] C. Herreid, N. Schiller, K. Herreid, and C. Wright, "Case study: a chat with the survey monkey: case studies and the flipped classroom," *Journal of College Science Teaching*, vol. 44, no. 1, 2014.
- [13] R. Pierce and J. Fox, "Vodcasts and active-learning exercises in a "flipped classroom" model of a renal pharmacotherapy module," *American Journal of Pharmaceutical Education*, vol. 76, no. 10, p. 196, 2012.
- [14] G. S. Mason, T. R. Shuman, and K. E. Cook, "Comparing the effectiveness of an inverted classroom to a traditional classroom in an upper-division engineering course," *IEEE Transactions on Education*, vol. 56, no. 4, pp. 430–435, 2013.
- [15] J. Cheng and X. Wang, "Artificial intelligence based on effectiveness of inverted classroom teaching of college sports," *Journal of Intelligent & Fuzzy Systems*, vol. 40, no. 2, pp. 3755–3765, 2021.
- [16] M. Persike, "Lean video production-towards ressource-efficient inverted classroom models," in *Global Learn Conference 2015 Association for the Advancement of Computing in Education (AACE)*, Berlin, 2015.
- [17] M. Boutell and C. Clifton, "SPLICE: self-paced learning in an inverted classroom environment," *Proceedings of the 42nd ACM technical symposium on computer science education*, 2011.
- [18] M. Mohammadi, M. R. Borujeni, A. Movahedi, and H. Salehi, "The effects of easy goals versus difficult goals on acquisition and retention of a sport skill in children with intellectual disability," *Journal of Intellectual Disabilities*, 2020.
- [19] R. E. Clumpner, "Physical education curriculum guidelines, grades K-12. Curriculum development," *Course Content*, p. 170, 1985.
- [20] J. Beard, *The effects of traditional and progressive physical education courses on adolescents' body compositions and student satisfaction [M.S. thesis]*, Walden University, 2012.
- [21] D. N. Daum and C. Buschner, "The status of high school online physical education in the United States," *Journal of Teaching in Physical Education*, vol. 31, no. 1, pp. 86–100, 2012.
- [22] A. M. Williams and A. Grant, "Training perceptual skill in sport," *International Journal of Sport Psychology*, vol. 30, no. 2, pp. 194–220, 1999.
- [23] M. Denis, "Visual imagery and the use of mental practice in the development of motor skills," *Canadian Journal of Applied Sport Sciences*, vol. 10, no. 4, pp. 45–165, 1985.
- [24] K. Ashwini, R. Ponuma, and R. Amutha, "Fine motor skills and cognitive development using virtual reality-based games in children," in *Handbook of Decision Support Systems for Neurological Disorders*, pp. 187–201, Elsevier, 2021.
- [25] J. Galecki, T. Zaradkiewicz, L. Kurpeta, and B. Bergier, "Activities of physical education teachers in the process of forming motor skills of the students," *Polish Journal of Sport and Tourism*, vol. 18, no. 1, pp. 33–38, 2011.
- [26] H. J. Syväoja, A. Kankaanpää, L. Joensuu et al., "The longitudinal associations of fitness and motor skills with academic achievement," *Medicine & Science in Sports & Exercise*, vol. 51, no. 10, pp. 2050–2057, 2019.
- [27] A. F. Machula and T. Department, "The development of motor skills and abilities of bullet shooting among cadets," vol. 1, no. 59, pp. 46–50, 2010.
- [28] H. Bin, H. F. Wang, and G. J. Xie, "Study on the influencing factors of crowdsourcing logistics under sharing economy," *Management Review*, vol. 31, no. 8, pp. 219–229, 2019.
- [29] R. Likert, "A technique for the measurement of attitudes," *Archives of Psychology*, vol. 22, no. 140, pp. 1–55, 1932.
- [30] F. Effendy, O. D. Kurniawati, and G. Priambada, "Factor affecting E-learning user acceptance: a case study of AULA," *Journal of Physics: Conference Series*, vol. 1783, article 012122, 2021.
- [31] F. Ali and R. Omar, "Determinants of customer experience and resulting satisfaction and revisit intentions: PLS-SEM approach towards Malaysian resort hotels," *Asia-Pacific Journal of Innovation in Hospitality and Tourism (APJIHT)*, vol. 3, no. 2, 2014.
- [32] L. Huang, G. Xie, J. Blenkinsopp, R. Huang, and H. Bin, "Crowdsourcing for sustainable urban logistics: exploring the factors influencing crowd workers' participative behavior," *Sustainability*, vol. 12, no. 8, p. 3091, 2020.
- [33] R. E. Buitrago, M. I. Barbosa Camargo, and F. Cala Vitery, "Emerging economies' institutional quality and international competitiveness: a PLS-SEM approach," *Mathematics*, vol. 9, no. 9, p. 928, 2021.
- [34] H. C. Kim, S. Kim, and Z. Y. Zhu, "Does foreign language proficiency help to enhance sustainable online brand community experiences? Modeling the predictors of movie information sharing behavior for young Chinese students staying in Korea," *Sustainability*, vol. 13, no. 11, p. 6113, 2021.
- [35] D. Falih Bannay, M. Jabbar Hadi, and A. Abdullah Amanah, "The impact of inclusive leadership behaviors on innovative workplace behavior with an emphasis on the mediating role of work engagement," *Problems and Perspectives in Management*, vol. 18, no. 3, pp. 479–491, 2020.

## Research Article

# An Improved Genetic Algorithm and Neural Network-Based Evaluation Model of Classroom Teaching Quality in Colleges and Universities

Huaying Zhang,<sup>1</sup> Bin Xiao,<sup>2</sup> Jinqiong Li,<sup>3</sup> and Min Hou <sup>1</sup>

<sup>1</sup>The Party Committee Organizes the Publicity Department, Hebei College of Industry and Technology, Shijiazhuang, 050091 Hebei, China

<sup>2</sup>Applied Technology Research and Development Center, Hebei College of Industry and Technology, Shijiazhuang, 050091 Hebei, China

<sup>3</sup>Hebei Women's Vocational College, Shijiazhuang, 050091 Hebei, China

Correspondence should be addressed to Min Hou; [houmin202102@163.com](mailto:houmin202102@163.com)

Received 17 June 2021; Revised 18 July 2021; Accepted 26 July 2021; Published 18 August 2021

Academic Editor: Shan Zhong

Copyright © 2021 Huaying Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Research on educational quality has gotten a lot of attention as the current higher education teaching reform continues to deepen and grow. The key to improving education quality is to improve teaching quality, and teacher evaluation is an important tool for doing so. As a result, educational management requires the development and refinement of a system for evaluating teaching quality. Traditional approaches to assessing teaching quality, on the other hand, are problematic due to their limitations. As a result, a scientific and reasonable model for evaluating the teaching quality of college undergraduate teachers must be developed. We present a unique model for evaluating the quality of classroom teaching in colleges and universities, which is based on improved genetic algorithms and neural networks. The basic idea is to use adaptive mutation genetic algorithms to refine the initial weights and thresholds of the BP neural network. The teaching quality evaluation findings were improved by improving the neural network's prediction accuracy and convergence speed, resulting in a more practical scheme for evaluating college and university teaching quality. We have conducted simulation experiments and comparative analysis, and the mean square error of the results of the proposed model is very low, which proves the effectiveness and superiority of the algorithm.

## 1. Introduction

The goal of teaching quality evaluation [1–4] is to promote teaching reform, improve teaching quality, reduce student burden [5–7], develop students' intelligence [8], and help students evaluate and solve problems. We must achieve the unity of ideology, science, and feasibility when evaluating the quality of teaching, and we must do so in an objective, fair, and rational manner, rather than subjectively guessing or mixing personal feelings [9, 10]. In colleges and universities, teaching quality is often assessed through four channels: student evaluation, expert evaluation, peer evaluation, and instructor self-evaluation, with the final evaluation results

synthesized. However, certain issues remain in the process of developing, utilizing, and evaluating the teaching quality assessment system's evaluation outcomes [3, 4], such as evaluation theory research [11], evaluation method usage, evaluation method update, and evaluation data analysis. These issues have a direct impact on educational institutions. In the future, quality assessment and knowledge extraction will be critical [12].

The indicators in the evaluation system generally involve teaching attitude [13], proficiency in teaching content, and basic teaching skills [14–16], etc. However, the comprehensive quality of teachers is not only reflected in the above aspects but also includes teachers' knowledge level, teaching

research ability, teaching design ability, and teachers' innovation ability, etc. But at present, these evaluation indexes which can fully reflect the comprehensive quality and personality of teachers are seldom involved in the evaluation system, so they should be fully considered in the establishment of the evaluation index system.

The indicators in the evaluation system [17] of the differences in the degree of influence of the evaluation results should be assigned different weights, but many colleges and universities still use the same weight method, or subjectively determined a weight distribution table to establish the evaluation system, and then use this evaluation system to evaluate. Therefore, the reasonable allocation of weight is the key step to perfect the evaluation system.

Teaching quality and teaching quality evaluation systems must be synchronized in colleges and universities; the current situation necessitates the construction of a teaching quality evaluation system; its positioning determines for teachers that teaching quality evaluation cannot be done solely through the theory of teaching evaluation; more attention must be paid to the cultivation of students. They can also address the needs of social development in the real world. As a result, assessing teaching quality is an important aspect of teaching management. We will use intelligent technology to make the evaluation of teaching quality more scientific and quantitative based on the above knowledge. At the same time, the study content of this paper serves as a diagnosis, feedback, and incentive, allowing for the early detection of difficulties in the teaching process and timely feedback to teachers in order to improve and improve teaching quality. Furthermore, scientific evaluation will apply appropriate pressure to teachers, motivating them to actively enhance the quality of their instruction and talent development.

The following are the main innovation points of this paper:

- (1) This research develops a model for evaluating college classroom teaching quality based on an improved genetic algorithm and a neural network, resulting in a novel way for evaluating college teaching quality. At the same time, it is expected to provide a valuable reference basis for the teaching management department to obtain scientific teaching quality evaluation work plans and programs, as well as provide reasonable judgments for the promotion and evaluation of teachers' professional titles, and make teaching management more scientific, institutionalized, and standardized
- (2) This research uses an adaptive mutation genetic technique to optimize the BP neural network's initial weights and thresholds. Because the BP neural network's initial weights and threshold value are so critical, utilizing a better genetic algorithm to optimize the initial weights and threshold value, reduce the BP neural network's training duration to satisfy the weight termination conditions and time threshold, and increase the neural network's teaching quality to the prediction accuracy assessment findings

## 2. Related Work

**2.1. Genetic Algorithm.** The basic idea of genetic algorithm [18–20] is to simulate the evolutionary process of the population, which is to conduct organized random information exchange and recombination for individuals [21, 22]. In the string structure of the previous generation, adaptive bits and segments are selected to recombine to generate a new generation of population, namely, “survival of the fittest.” As an additional addition, occasionally, new bits and segments are added to the string structure to replace the original ones, known as “mutations.” After the three genetic operations of selection, crossover, and mutation, the population is constantly updated, the population's good degree is constantly enhanced, and the global optimal solution is approached. The process of the standard genetic algorithm is shown in Figure 1.

**2.2. BP Network.** Compared with the other intelligent models [23], the back propagation neural network, also known as the feedforward neural network [24–26], is a very simple prediction model, and it has a three-layer feedforward hierarchical network with input, hidden, and output layers [27, 28]. The topology of the three-layer feedforward neural network is shown in Figure 2. When a set of input modes is presented to the network, the BP network will learn the set of input modes in the following order: first, the hidden layer unit receives the input mode from the input layer. An input mode is generated and delivered to the output layer after the hidden layer unit processes the input mode layer by layer. Forward propagation is the term for this phenomenon. The output findings are then compared to the predicted values. If the expected values are not met, error back propagation is used. Error signals are decreased by altering the connection weights of neurons at each layer, and the error is returned along the original path. As part of a “memory training” process, forward propagation and back propagation alternate. The system repeats these two stages, learning until the difference between the output value and the expected value is within a certain range, at which point the system stops learning. The fresh sample is now fed into the trained network, and the associated output value is calculated.

For the hidden layer, there are

$$y_j = f(\text{net}_j), j = 1, 2, \dots, m, \quad (1)$$

$$\text{net}_j = \sum_{i=0}^n v_{ij}x_i, j = 1, 2, \dots, m, \quad (2)$$

where  $f(x)$  is

$$f(x) = \frac{1}{1 + e^{-x}}, \quad (3)$$

$$f'(x) = f(x)[1 - f(x)]. \quad (4)$$

When the network output is not equal to the expected

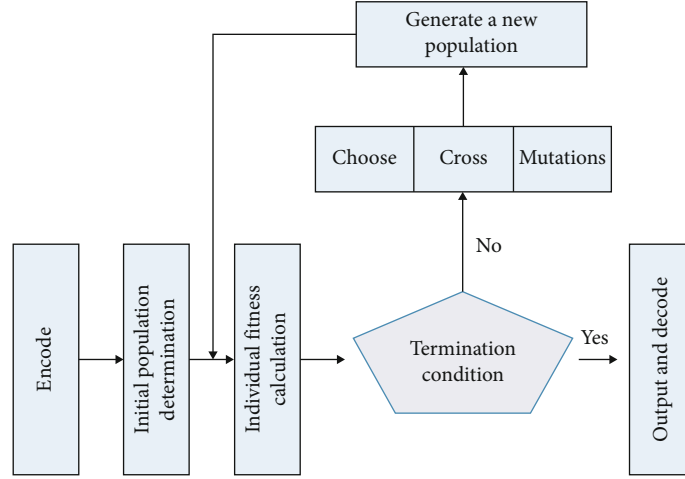


FIGURE 1: Schematic diagram of genetic algorithm.

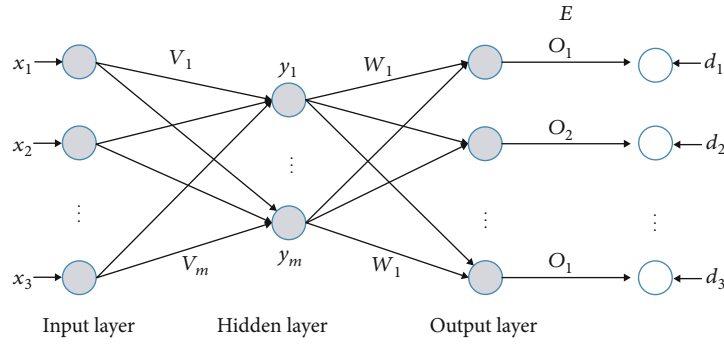


FIGURE 2: Schematic diagram of genetic algorithm.

output, the output error  $E$  exists, which is defined as follows:

$$E = \frac{1}{2} (d - o)^2 = \frac{1}{2} \sum_{k=1}^1 (d_k - o_k)^2. \quad (5)$$

Expand the above error definition to the hidden layer, then:

$$E = \frac{1}{2} \sum_{k=1}^1 [d_k - f(\text{net}_k)]^2 = \frac{1}{2} \sum_{k=1}^1 \left[ d_k - f \left( \sum_{j=0}^m W_{jk} Y_j \right) \right]^2. \quad (6)$$

It can be seen from the above formula that the network input error is a function of the weights  $w_{jk}$  and  $v_{ij}$  of each layer, so adjusting the weights can change the error  $E$ . Obviously, the principle of adjusting the weight is to continuously reduce the error, so the adjustment of the weight should be proportional to the negative gradient of the error, that is,

$$\Delta w_{jk} = -\eta \frac{\partial E}{\partial w_{jk}}, \quad (7)$$

$$\Delta v_{ij} = -\eta \frac{\partial E}{\partial v_{ij}}. \quad (8)$$

The negative sign in the equation represents the gradient descent, and the constant  $\eta \in (0, 1)$  represents the proportional coefficient, which reflects the learning rate during training. It can be seen that the BP algorithm belongs to the  $\delta$  learning rule class.

### 3. Methodology

**3.1. Adaptive Mutation Genetic Algorithm.** The calculation equation of adaptive mutation probability  $P$  is as follows:

$$P = \frac{(P_1 + P_2)}{2} = \frac{((P_0(P_0 - P_{\min}) * m/M) + (P_0 * \max F(X_k)/\bar{F}))}{2}, \quad (9)$$

where  $M$  is the maximum evolutionary algebra,  $m$  is the current evolutionary algebra,  $P_1$  is inversely proportional to the evolutionary algebra,  $P_2$  is inversely proportional to the average fitness value,  $P_0$  is the assumed initial mutation probability,  $P_{\min}$  is the minimum value of the mutation probability range, and  $\bar{F}$  is the average fitness value of the current group, which is the maximum fitness value of the current group.

### 3.2. Adaptive Mutation Genetic-BP Model

**3.2.1. Network Construction.** All continuous functions can be mapped using a feedforward neural network with a single hidden layer. Two hidden layers are only required for learning discontinuous functions. As a result, a multilayer feedforward neural network requires no more than two hidden layers. In general, the initial step in creating a multilayer feedforward neural network is to create a hidden layer. If the hidden layer has a big enough number of nodes and the network performance does not improve, the training cost will rise as the number of hidden layers grows. As a result, this post tries to employ a hidden layer initially. Because the input layer gets data from the outside, the number of nodes is determined by the size of the problem's input vector. The transfer function used by the input layer is generally a linear function, that is,  $f(x) = x$ . The trial-and-error method is one of the methods to determine the number of hidden layer nodes. After this procedure has found the initial value, experiments can be carried out by raising the number from small to small and analyzing the outcomes to identify the best number. The trial and error approach has three ways to determine the initial value, and the calculation equation is as follows:

$$m = \sqrt{n + l} + \alpha, \quad (10)$$

$$m = \log 2^n. \quad (11)$$

**3.2.2. Adaptive Mutation Genetic Algorithm.** A genetic algorithm's goal is to find network weights and thresholds that minimize the network's sum of squared errors over all evolutionary generations, while the fitness function evolves in the direction of increasing its value, making the fitness function the inverse of each individual learning error. The following are the learning error and fitness function calculation equations:

$$E = \frac{\sum_{k=1}^p \sum_{j=1}^l (y_j^k - o_j^k)^2}{2}, \quad (12)$$

$$\text{fitness} = \frac{1}{E}, \quad (13)$$

where  $E$  is the learning error,  $p$  is the number of training samples that is 2000 sets of evaluation data,  $l$  is the number of output nodes 1, and  $y_j^k - o_j^k$  is the error of the  $k$ -th sample relative to the  $j$ -th output node.

The mutation operation is the process by which the genes of some people in a population mutate with a certain probability. The adaptive mutation probability mutation operation is used in the model. Although bad individual shapes will appear to some extent, the genetic operation method of mutation will retain some favorable mutations, increase the diversity of the genetic algorithm population, and cause it to jump out of the

local optimal solution in time, search for the global optimal solution, and avoid premature phenomena.

**3.3. Teaching Quality Evaluation Model.** First, by analyzing existing problems in teaching quality evaluation, we can improve them and establish a more complete and more appropriate index system. Collect teaching quality evaluation sample data, select evaluation indicators according to the teaching characteristics of teachers, and divide the collected teaching quality evaluation data into training samples and test samples. Second, determine the learning rate, the number of hidden layer neurons, the maximum number of iterations, the minimum error accuracy, the transfer function, the number of training, and other parameters of the BP neural network method. By inputting samples into the evaluation model, iterative training is continuously carried out until the triggering algorithm stops. Then, for teaching quality evaluation, enter the test sample to see if the training impact of the enhanced genetic algorithm-optimized BP neural network model fulfills the requirements. Enter the next phase if the prediction result meets the stop criteria; otherwise, return to the previous stage and retrain the network. Finally, to obtain the teaching quality evaluation result, input the sample into the teaching quality evaluation model.

## 4. Experiments and Results

**4.1. Experimental Environment.** The experimental system software environment used in this article is shown in Table 1.

When updating parameters,  $lr$  means the learning rate is 0.0001. The experiments with all the algorithms were performed on a computer equipped with a single NVIDIA GTX1080 GPU (8 GB).

**4.2. Data Collection.** Teaching quality evaluation consists of four parts: leader evaluation, expert evaluation, peer evaluation, and student evaluation. The methods of obtaining teaching quality evaluation data are as follows: (1) *Leadership Evaluation.* Take random lectures and evaluate the teacher's teaching and student learning. (2) *Expert Evaluation.* The Academic Affairs Office and each college will determine the evaluation courses, respectively, and the expert group will conduct inspection courses. (3) *Peer Evaluation.* Organize experienced teachers to evaluate peer teachers and adopt the methods of listening, evaluating, and discussing lectures to improve the teaching strategies and methods of the assessed teachers and improve their teaching ability. (4) *Student Evaluation.* Every semester, students evaluate the teaching quality of their own class teachers. Teachers' teaching quality evaluation is usually arranged in the middle of the semester and before the final exams in each semester. Our data set consists of 2 data from different universities, named Data1 and Data2, respectively.

**4.3. Evaluation Index.** We use the mean square error to evaluate the proposed algorithm, and its calculation equation is as follows:

$$\text{MSE} = \frac{1}{mp} \sum_{p=1}^p \sum_{j=1}^m (y^{\wedge}_{pj} - y_{pj})^2, \quad (14)$$

TABLE 1: Parameters of the experimental environment.

Type	Parameter
OS	Windows 10
CPU	Intel Core I5
RAM	8.00 GB
Development tools	Pycahrn

TABLE 2: Comparison of the results of the experiment on Data1.

Sample number	BP		Ours	
	Prediction	MSE	Prediction	MSE
1	11.02	7.25%	10.23	5.22%
2	15.62	6.12%	16.21	4.36%
3	14.20	5.26%	14.69	4.12%
4	11.98	6.21%	12.01	5.36%
5	12.17	4.31%	13.01	3.33%
6	12.36	6.25%	11.98	4.11%

TABLE 3: Comparison of the results of the experiment on Data2.

Sample number	BP		Ours	
	Prediction	MSE	Prediction	MSE
1	10.03	7.15%	11.24	5.11%
2	16.11	6.31%	15.99	5.08%
3	13.11	5.11%	13.89	3.93%
4	10.19	6.32%	11.88	4.99%
5	12.33	3.89%	11.09	4.11%
6	12.12	5.91%	11.23	3.99%

where  $m$  is the number of output nodes,  $p$  is the number of training samples,  $\hat{y}_{pj}$  is the expected output value of the network, and  $y_{pj}$  is the actual output value of the network. Compared with the standard BP neural network error, the reduction of the output error of other samples will not directly lead to the increase of iteration times after the weight modification. The setting of cumulative error is to reduce the global error of the whole training set, rather than the error of a specific small sample. Therefore, the mean square error is more reasonable than the cumulative error.

**4.4. Experimental Results.** The results of the training of the teaching quality evaluation model based on the proposed method and the traditional BP method are compared, and the results are shown in Tables 2 and 3.

Compared with the prediction results of the traditional BP algorithm for 6 groups of samples, the error between the output value of the teaching effect measured by our method and the real value is relatively small. To see if BP neural network has better approximation ability and more accurate prediction effect based on improved genetic algorithm and neural network of institutions of higher learning in classroom teaching quality evaluation model of ability to predict on the teaching quality evaluation prediction, thus, more scientific and accurate evaluation of college teaching quality and teaching effect showed the effectiveness of the

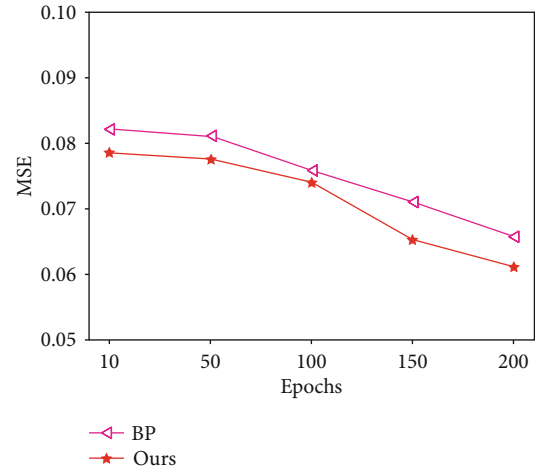


FIGURE 3: Mean square error of the proposed method and BP on Data1.

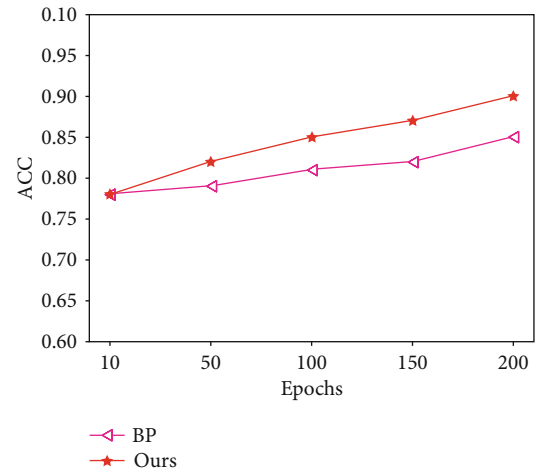


FIGURE 4: Acc of the proposed method and BP on Data1.

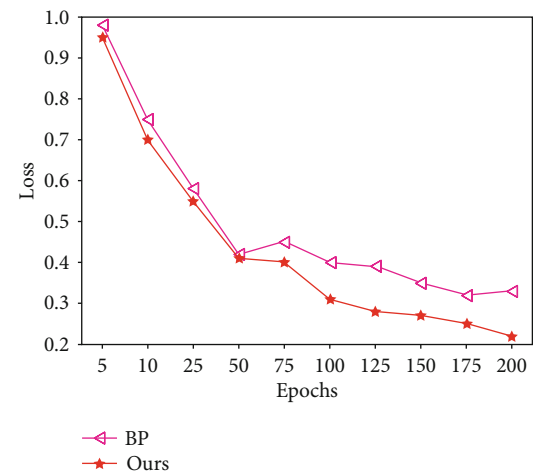


FIGURE 5: Loss of the proposed method and BP on Data1.

model. Figures 3 and 4 also show the MSE and ACC comparison curve during the training process, and Figure 5 shows the training loss curve.

## 5. Conclusion

In recent years, improving the quality of higher education teaching has been a top concern, and teacher evaluation is an important measure of educational and instructional quality. As a result, educational administration requires the development and refining of a system for measuring teaching quality. Traditional teaching quality evaluation approaches, on the other hand, have been rendered ineffective due to their limitations. As a result, a scientific and fair teaching quality evaluation model must be developed to assess the teaching quality of college undergraduate teachers. We present a unique approach for measuring the quality of classroom teaching in colleges and universities, which is based on improved genetic algorithms and neural networks. The basic idea is to use adaptive mutation genetic algorithms to modify the initial weights and thresholds of the BP neural network. Because the BP neural network's initial weight and threshold are so crucial, the improved genetic algorithm is used to optimize the BP neural network's initial weight and threshold in order to reduce the time it takes for the BP neural network to find the weight and threshold that meets the training termination condition. Improving the neural network's prediction accuracy and convergence speed to the teaching quality evaluation findings resulted in a more practical scheme for evaluating college and university teaching quality.

## Data Availability

The data used to support the findings of this study are included within the article.

## Conflicts of Interest

All the authors do not have any possible conflicts of interest.

## References

- [1] P. Spooren, D. Mortelmans, and J. Denekens, "Student evaluation of teaching quality in higher education: development of an instrument based on 10 Likert-scales," *Assessment & Evaluation in Higher Education*, vol. 32, no. 6, pp. 667–679, 2007.
- [2] J. Douglas and A. Douglas, "Evaluating teaching quality," *Quality in Higher Education*, vol. 12, no. 1, pp. 3–13, 2006.
- [3] B. Dunrong and M. Fan, "On student evaluation of teaching and improvement of the teaching quality assurance system at higher education institutions," *Chinese Education & Society*, vol. 42, no. 2, pp. 100–115, 2009.
- [4] V. J. Maslow and C. J. Kelley, "Does evaluation advance teaching practice? The effects of performance evaluation on teaching quality and system change in large diverse high schools," *Journal of school leadership*, vol. 22, no. 3, pp. 600–632, 2012.
- [5] O. Luaces, J. Díez, and A. Bahamonde, "A peer assessment method to provide feedback, consistent grading and reduce students' burden in massive teaching settings," *Computers & Education*, vol. 126, pp. 283–295, 2018.
- [6] F. Ullah, A. Bajahzar, H. Aldabbas et al., "An e-assessment methodology based on artificial intelligence techniques to determine students language quality and programming assignments plagiarism," *Intelligent Automation & Soft Computing*, vol. 26, no. 1, pp. 169–180, 2020.
- [7] P. Tang, Y. Wang, and N. Shen, "Prediction of college students' physical fitness based on K-means clustering and SVR," *Computer Systems Science and Engineering*, vol. 35, no. 4, pp. 237–246, 2020.
- [8] J. J. Fitzpatrick, "Helping nursing students develop and expand their emotional intelligence," *Nursing Education Perspectives*, vol. 37, no. 3, p. 124, 2016.
- [9] J. Zhang, "Personalised product recommendation model based on user interest," *Computer Systems Science and Engineering*, vol. 34, no. 4, pp. 231–236, 2019.
- [10] J. T. Sears, "Educators, homosexuality, and homosexual students: are personal feelings related to professional beliefs?," *Journal of Homosexuality*, vol. 22, no. 3-4, pp. 29–80, 1992.
- [11] M. C. Alkin and C. A. Christie, "An evaluation theory tree," *Evaluation roots: Tracing theorists views and influences*, vol. 2, no. 19, pp. 12–65, 2004.
- [12] J. Luan, *Data Mining and Knowledge Management in Higher Education-Potential Applications*, 2002, <https://eric.ed.gov/?id=ED474143>.
- [13] S. Mahony and E. Pierazzo, *Teaching Skills or Teaching Methodology?*, OpenBook Publishers, 2012.
- [14] G. Gibbs and M. Coffey, "The impact of training of university teachers on their teaching skills, their approach to teaching and the approach to learning of their students," *Active Learning in Higher Education*, vol. 5, no. 1, pp. 87–100, 2004.
- [15] J. M. Braxton, N. J. Bray, and J. B. Berger, "Faculty teaching skills and their influence on the college student departure process," *Journal of College Student Development*, vol. 41, no. 2, 2000.
- [16] J. Johnston, J. Halocha, and M. Chater, *Developing Teaching Skills in the Primary School*, McGraw-hill education (UK), 2007.
- [17] C. Wang and N. Gao, "Study of the plural evaluation indicators system and evaluation model in physical education teaching in institution of higher learning," *Journal of Computational and Theoretical Nanoscience*, vol. 14, no. 1, pp. 177–181, 2017.
- [18] A.-A. Adebayo, S. Misra, L. Fernandez-Sanz, A.-A. Olusola, and A. R. Edun, "Genetic algorithm and tabu search memory with course sandwiching (gats\_cs) for university examination time-tabling," *Intelligent Automation & Soft Computing*, vol. 26, no. 3, pp. 385–396, 2020.
- [19] F. Chou, W. Ho, and C. Chen, "Niche genetic algorithm for solving multiplicity problems in genetic association studies," *Intelligent Automation & Soft Computing*, vol. 26, no. 3, pp. 501–512, 2020.
- [20] A. Alhroob, W. Alzyadat, A. T. Imam, and G. M. Jaradat, "The genetic algorithm and binary search technique in the program path coverage for improving software testing using big data," *Intelligent Automation & Soft Computing*, vol. 26, no. 4, pp. 725–733, 2020.
- [21] A. Y. Hamed, M. H. Alkinani, and M. R. Hassan, "A genetic algorithm optimization for multi-objective multicast routing," *Intelligent Automation & Soft Computing*, vol. 26, no. 4, pp. 1201–1216, 2020.
- [22] J. Ye, "PID tuning method using single-valued neutrosophic cosine measure and genetic algorithm," *Intelligent Automation & Soft Computing*, vol. 25, no. 1, pp. 15–23, 2019.
- [23] Y. Jiang, X. Gu, D. Wu, W. Hang, J. Xue, and S. Qiu, "A novel negative-transfer-resistant fuzzy clustering model with a shared cross-domain transfer latent space and its application to brain CT image segmentation," *IEEE/ACM Transactions*



- on Computational Biology and Bioinformatics*, vol. 18, no. 1, pp. 40–52, 2021.
- [24] X. Ning, Y. Wang, W. Tian, L. Liu, and W. Cai, “A biomimetic covering learning method based on principle of homology continuity,” *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 1, pp. 9–16, 2021.
- [25] X. Ning, W. Li, and J. Xu, “The principle of homology continuity and geometrical covering learning for pattern recognition,” *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 32, no. 12, p. 1850042, 2018.
- [26] Y. Tong, L. Yu, S. Li, J. Liu, H. Qin, and W. Li, “Polynomial fitting algorithm based on neural network,” *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 1, pp. 32–39, 2021.
- [27] X. Ning, K. Gong, W. Li, L. Zhang, X. Bai, and S. Tian, “Feature refinement and filter network for person re-identification,” *IEEE Transactions on Circuits and Systems for Video Technology*, p. 1, 2020.
- [28] R. Liu, X. Ning, W. Cai, and G. Li, “Multiscale dense cross-attention mechanism with covariance pooling for hyperspectral image scene classification,” *Mobile Information Systems*, vol. 2021, 15 pages, 2021.

## Research Article

# Convolutional Neural Network-Assisted Strategies for Improving Teaching Quality of College English Flipped Class

Tiankun Liu 

*School of Foreign Languages, Hulunbuir University, Inner Mongolia Hulunbuir, 021008, China*

Correspondence should be addressed to Tiankun Liu; [liutiankun@hlbec.edu.cn](mailto:liutiankun@hlbec.edu.cn)

Received 13 July 2021; Revised 3 August 2021; Accepted 6 August 2021; Published 15 August 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Tiankun Liu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The “flipped classroom” teaching paradigm not only follows the cognitive rules of the learners, but it also subverts and reverses the standard classroom teaching process. Problem-oriented, teacher-led, student-centered, and mixed teaching approaches are the key teaching methods in the flipped classroom teaching model, which focuses on students’ procedural knowledge acquisition and critical thinking training. There are a lot of studies on the specific practice path of the “flipped classroom” teaching style right now, but there are not many on the learning involvement of college English students in this approach. According to studies, the level of student participation in classroom learning is the most important factor limiting the efficiency of teaching. The lack of research in this subject greatly limits the “flipped classroom” teaching model’s ability to improve college English classroom teaching quality. The degree of engagement between teachers and students, the enthusiasm of students in class, and the competence of teachers to educate are all reflected in student conduct in the classroom. Understanding and evaluating the behaviors and activities of students in the classroom are helpful in determining the state of students in the classroom, as well as improving the flipped classroom teaching technique and quality. As a result, the convolutional neural network is used to recognize student behavior in the classroom. The loss function of VGG-16 has been enhanced, the distance inside the class has been lowered, the distance between classes has been increased, and the recognition accuracy has improved. Accurate recognition of classroom behavior is beneficial in developing methods to improve teaching quality.

## 1. Introduction

The flipped classroom teaching model [1–3] has grown into a magnificent landscape of education and teaching reform [4] as a new teaching model [5–7]. Classroom teaching content in traditional classrooms relies on one-way teaching and transmission of book knowledge. The time for students to acquire internalized knowledge in the classroom is very restricted due to classroom teaching time constraints, and the influence of classroom teaching on the development of students’ critical thinking capacity is not ideal. The flipped classroom is learner-centered and problem-oriented, and the classroom flipping of declarative and procedural knowledge allows teachers to devote more time in the classroom to students answering questions and deepening their knowledge understanding. Simultaneously, students bring problems from preclass research into the classroom [8–10], resulting in more focused and successful classroom discussions. In

traditional classrooms, students’ learning is also easy to change from passively listening to lectures or transcribing teachers’ classroom teaching notes. Constructiveness (based on learner experience), structurality (knowledge structure development), criticality (critical assessment of knowledge and viewpoints), comprehension (memory learning that leads to understanding), and reflectiveness are all factors that contribute to efficiency (continuously in the learning process). The phenomena of collective quiet among students in classroom teaching have been overcome by in-depth learning defined by reflection and monitoring.

English acquisition [11, 12], in the final analysis, is to achieve better communication between people. Silence in the process of classroom English acquisition will inevitably lead to obstacles to English learning. At present, university classroom teaching can be roughly divided into “five levels,” namely, silence, answer, dialogue, critical, and debate. Teaching practice shows that the use of the flipped classroom

teaching model in college English classroom teaching can well promote the improvement of foreign language learners' various language skills, greatly improve the quality of college English classroom teaching in colleges and universities, and help deepen the reform of college English education and teaching. The flipped classroom teaching model takes students as the center of classroom teaching, allows students to participate in the process of teaching activities, and greatly improves students' participation and enthusiasm for course learning [13]. In order to explore the impact of flipped learning on English learners' second-language speaking, second-language listening, and participation and participation in curriculum materials and activities outside the classroom, some scholars divided 67 English university freshmen into three groups: structured flipped learning group, semistructured flipped learning group, and the traditional learning group; the research results show that flipped learning helps to improve the oral and listening skills of English learners and enables them to participate more in extracurricular activities. It can be seen that the flipped classroom teaching model can indeed successfully bring the traditional Chinese college English classroom out of the "silent" quagmire and make the classroom teaching model move towards the realm of "dialogue," "questioning," and "debating."

However, the current foreign scholars' research on the flipped classroom teaching model mainly focuses on theoretical research such as teaching model construction, teaching practice exploration and application [14], teaching implementation methods [15, 16], comparison studies with traditional classroom teaching models, and empirical research on the effects of teaching practice after the model is applied. However, most of the academic research on the application of the flipped classroom teaching model of college English focuses on the teacher teaching mode, student learning mode, teacher teaching ability (teaching design, team building, and information literacy), and course assessment methods of the flipped classroom teaching model. However, there are relatively few studies on the degree of student participation in classroom learning behind the phenomenon of classroom activity in the literature. The monitoring and promotion of students' participation [17] in classroom learning belong to the category of teaching quality monitoring. A high degree of student participation in classroom learning is an indispensable prerequisite for any teaching model to ensure its teaching quality. If there is only an increase in the "temperature" of the atmosphere in the classroom, but no deep participation of students in the true sense of learning, then the implementation of the flipped classroom teaching model will greatly lose its true meaning of education.

Since student conduct in the classroom reflects the level of interaction between teachers and students [18, 19], students' interest in class, and teachers' competence to educate, understanding and evaluating the actions and activities of students in the classroom are beneficial to understanding the status of students in class, as well as improving flipped classroom teaching methods and improving the quality of flipped classroom teaching. As a result, convolutional neural networks [20–23] can be used to recognize student behavior in classes [24, 25].

The main contributions of this paper are as follows:

- (1) This paper proposes a teaching quality promotion model for college English flipped classroom based on the assistance of convolutional neural network, which can improve the teaching method of flipped classroom and improve the quality of flipped classroom teaching
- (2) Aiming at the characteristics of similar faces, the loss function of VGG-16 is improved to reduce the intra-class distance and increase the interclass distance. The improved VGG-16 network improves the accuracy of emotion recognition in classroom students

## 2. Background

At present, the research on class behavior recognition is mainly reflected in the behavior state of students in class, such as raising hands, standing up, and sleeping. There are also some studies that reflect students' abnormal behavior in the examination room. These studies can also be classified as classroom behavior recognition because they are behavior recognition in a classroom setting. The standing recognition algorithm based on the region of interest uses the characteristic that the standing behavior occurs in the upper half of the image, crops the upper half of the image into the region of interest, and realizes standing recognition under different backgrounds through different threshold segmentation algorithms [26, 27]. The student behavior recognition algorithm based on the ResNet network [28] first trains on the ImageNet data set, then uses transfer learning to apply the ResNet network to student behavior recognition, and realizes the recognition of student behaviors such as looking left and right, raising hands, standing, and sleeping. Based on Faster R-CNN to identify the behavior state of students in the classroom, the YOLOV3 algorithm is used to extract the behavior sequence of the students, and finally, the behavior of students is classified through the ResNet network to realize the recognition of the behavior state of the students in the classroom. Based on the gradient histogram and the equivalent local binary mode histogram, the students' head-and-shoulders characteristics are merged, and the support vector machine is used to train the classifier to achieve target detection on the experimental data set. An algorithm based on sparse reconstruction is also proposed. Based on the 3D convolutional neural network, the time dimension is added to the two-dimensional convolutional neural network, which can better learn the time domain features. Simonyan and Zisserman [29] proposed a dual-stream CNN algorithm for human behavior recognition. This algorithm trains two CNN classifiers, one CNN mainly extracts optical flow features, and the other CNN extracts RGB image information, and finally the two classifiers fusion of features. Feichtenhofer et al. [30] proposed a new spatiotemporal structure based on the dual-stream architecture, which has a new convolutional fusion layer and a spatial fusion layer, which can better extract human behavior characteristics.

The detection and analysis of student behavior in the classroom scene can rapidly and efficiently identify the student's learning status and the teacher's teaching quality,

allowing for focused teaching technique changes and increased student learning efficiency. In general, students' behaviors in class include the following: paying attention in class, raising hands, standing up, napping, and using cell phones. Neural networks can be used to monitor and analyze student behavior [31, 32].

### 3. Methodology

*3.1. Flipped Classroom Theory.* The essence of the flipped classroom is to return the dominance of learning to the students, lead the students' subjective teaching methods, and have the energy to create education for the future. And "flipped classroom as a teaching concept and teaching model is affecting and changing traditional classroom teaching." It uses Internet technology and information technology to break through the boundaries of traditional classrooms, expand the time and space of classroom teaching, and optimize. The learning process of students enhances students' learning ability, realizes the deep integration of artificial intelligence and curriculum teaching, and promotes students' deep learning. At the same time, the flipped classroom model, as a part of the education reform movement, will completely subvert the traditional printing-based classroom teaching structure and teaching process and trigger a series of changes in the role of teachers, curriculum models, and management models.

Autonomous learning theory, cooperative learning theory, and mastering learning theory are three common theories in flipped classrooms.

Unlike the typical accepted learning approach, autonomous learning places a greater emphasis on pupils' ability to study independently. Students achieve their learning goals through independent analysis, investigation, practice, and invention as the main body of learning. Autonomous learning theory is founded on the principle of inquiry learning, which involves presenting students with a situation in which they must do their own research, solve problems, and gain expertise in the topic. Students in the classroom are given additional opportunity to "experience and interact" with knowledge as part of discovery learning. Traditional education and teaching approaches emphasize passive acceptance as a learning mode. Autonomous learning needs instructors to prioritize school instruction and complement it with required, scientific, and reasonable family and social education, so that children can learn to seek information, live, and survive through autonomous learning and have the ability to adapt to modern society. Encourage pupils to have a more and deeper understanding of knowledge in order to build the required abilities and basic attributes to continuously support their own development.

Cooperative learning is aimed at organizing classroom activities to promote academic and social learning experiences. Cooperative learning is not just about dividing students into small groups, but has been described as "building positive interdependence." In cooperative learning, students must work together in small groups and work together to achieve learning goals. Unlike individual learning, which can be competitive in nature, cooperative learning

students can leverage each other's resources and skills, seek information from each other, evaluate each other's ideas, monitor each other's work, and so on. In addition, the role of the teacher has also changed, from teaching knowledge to facilitating students' learning and from teaching to guiding. Compared with students in individual or competitive learning environments, students in cooperative learning environments achieved more and better reasoning, higher self-esteem and, for example, more social support when a group of classmates completed many learning tasks together.

According to the mastery of learning theory, education should be centered on the amount of time it takes for different students to acquire the same information and achieve the same level of mastery. To put it another way, all pupils can attain the same degree of knowledge as long as they have adequate time. The difference in a student's learning ability is only connected to the amount of time it takes him to master the knowledge, not whether or not he can master the topic. Traditional teaching focuses on the disparities in students' skills, and students' learning time and teaching techniques are essentially the same. Learning theory mastery contrasts sharply with typical teaching methods. Education is no longer just to allow a small group of students to fully learn what the school teaches, but to care about the development of each student and provide all students with the necessary knowledge and skills. In mastering learning, the responsibility of learning has changed. The failure of the students is more due to the factors of guidance, not necessarily the lack of ability of the learners. Therefore, in a mastered learning environment, the challenge becomes to provide students with enough time and appropriate teaching strategies so that all students can reach the same level of learning.

*3.2. Improved VGG-16 Network.* According to the size of the convolution kernel and the depth of the network, the VGG network model is divided into six network structures, including A, A-LRN, B, C, D, and E. VGG-16 is a class D network structure. VGG-16 consists of 13 convolution layers and 3 full connection layers. Each convolution layer contains the pooling layer and activation function, and the size of the convolution kernel is  $3 \times 3$ . The network diagram of VGG-16 is shown in Figure 1. The size of the fixed input image of each VGG-16 model is  $224 \times 224 \times 3$ . The first and second layers are  $64 \times 3 \times 3$  convolution kernels with a step size of 1, and the maximum pooling operation is adopted. The third layer and the fourth layer are  $128 \times 3 \times 3$  convolution kernels, which also adopt the maximum pooling operation. The size of the convolution kernel in the fifth, sixth, and seventh layers is 33, and the number of convolution kernels is 256. The remaining six convolutional layers have 512 convolution cores, and the maximum pooling operation is performed every three layers. Finally, a comprehensive connection layer with three levels is introduced. The first two layers of the entire connection layer include 4096 neurons, whereas the last layer contains 1000 neurons. Finally, the predicted value is determined using the softmax function's categorization.

VGG-16 was a simple and easy-to-train deep network structure at the time, and it produced good results in image recognition. VGG-16 is separated into various blocks, as

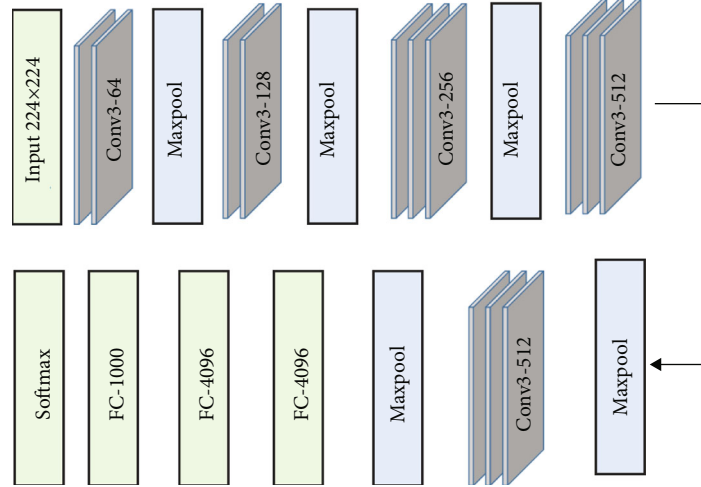


FIGURE 1: VGG-16 network structure diagram.

indicated in Figure 1. There are numerous convolutional layers and a pooling layer in each block. The convolutional layer's number of channels remains constant in the same block. After reaching 512, the number of convolution kernels doubles, doubling the number of channels, and the total number of channels remains unaltered. Despite the fact that many other network topologies have since been presented, VGG-16 remains a popular convolutional neural network model.

The classic VGG-16 network contains three fully connected layers, which leads to too many VGG-16 network parameters and slower training speed. This paper uses a face recognition algorithm based on the improved VGG-16 network to achieve the purpose of student face recognition in the classroom scene. By improving the classic VGG-16 network structure, it is more suitable for face recognition scenes; for face recognition scenes, the loss function is improved. In the task of student face recognition, the student's face is usually centered in the center of the image, and I hope that traits like eyes are learnt in various areas across the image. As a result, the last layer of VGG-16, as well as the third-to-last convolutional layer, is changed to a local convolutional layer in this article, allowing various parts of the image to learn completely separate features. At the same time, the VGG-16 network's fully connected layer is lowered by two layers, and the last layer's pooling layer is enhanced to an average pooling layer. 13 convolutional layers, 5 pooling layers, and 1 fully connected layer make up the upgraded VGG-16 network.

**3.3. Improved Loss Function.** I propose using center loss to build a mixed loss function that will increase the model's discriminative ability as well as its generalization capacity. Softmax loss can categorize images, and the central loss function can increase the distance between classes while decreasing the distance inside the class, as well as improve the model's accuracy. The center loss function's calculation equation is as follows:

$$L_C = \frac{1}{2} \sum_{i=1}^m \|x_i - C_{y_i}\|_2^2. \quad (1)$$

And the calculation equation of the softmax loss function is as follows:

$$L_S = - \sum_{i=1}^m \log \frac{\exp(w_{y_i}^T x_i + b_{y_i})}{\sum_{j=1}^n \exp(w_j^T x_i + b_j)}. \quad (2)$$

The calculation equation of the improved mixed loss function is as follows:

$$L = L_S + \lambda L_C, \quad (3)$$

where  $\lambda$  is the weight of the central loss function.

**3.4. Classroom Student Behavior Recognition.** The teaching quality promotion strategy of the English flipped classroom proposed by us is divided into two parts. First, I use the improved VGG-16 for student face recognition and behavior recognition. Second, I evaluate the recognition results and give each student developed teaching strategies and plans.

**3.4.1. Classroom Student Behavior Recognition.** The improved VGG-16 network structure for student face and behavior recognition is shown in Figure 2.

As shown in Figure 2, what I propose is a convolutional neural network architecture with twin parallel branches. The upper branch is mainly used for student facial emotion recognition, and the lower branch is mainly used for subject behavior recognition. And through the optimized softmax loss and center loss, respectively, the proposed algorithm has a higher accuracy rate.

**3.4.2. Teaching Quality Promotion Strategy.** Figure 3 shows the teaching quality promotion strategy I proposed in flipped classroom. First of all, I evaluate the students' face recognition and behavior recognition, respectively. The evaluation is based on the score of the control group.

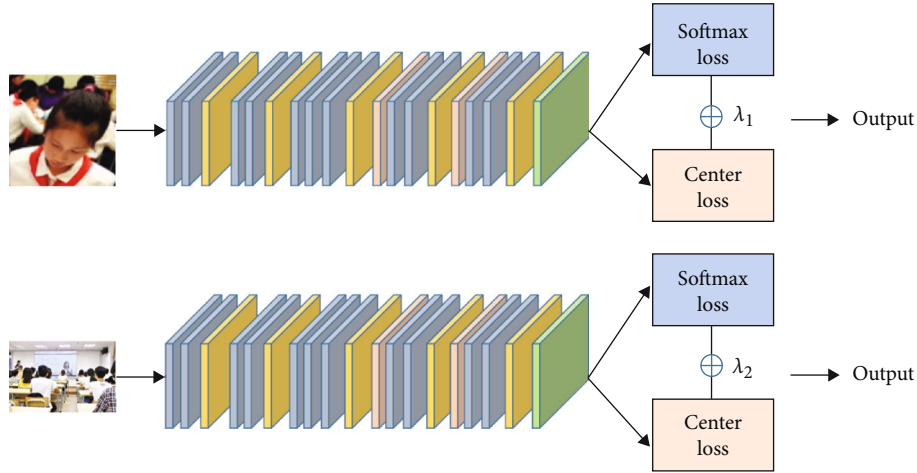


FIGURE 2: Schematic diagram of student face and behavior recognition architecture.

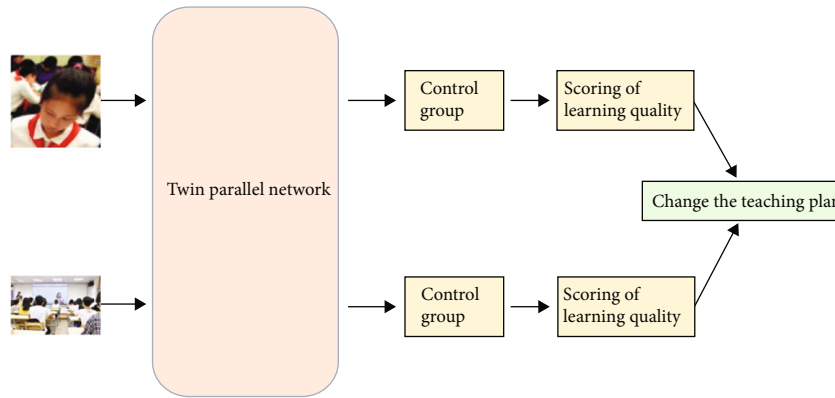


FIGURE 3: Schematic diagram of teaching quality promotion strategy framework.

## 4. Experiments

**4.1. Environment Configuration.** The experimental hardware environment in this paper is dual-channel Xeon E5 2678 V3 processor, the main CPU is 2.5 GHz, a total of  $12 * 2$  cores, the memory is 16 G, and the graphics card is NVIDIA RTX2060 Super (8 G video memory). The software environment of this experiment is Ubuntu 18.04 LTS operating system, the programming language is Python, the deep learning framework is TensorFlow and Caffe, and the GPU acceleration is carried out using CUDA10.0+ CUDNN7.6.4. The hyperparameter settings are shown in Table 1, and we divide the data set into a training set and a test set, 70% and 30%, respectively, and the batch size is 100.

**4.2. Data Sets.** The data set used to train face recognition models in this study was Labeled Faces in the Wild (LFW). Face identification in an uncontrolled context is studied using the LFW data set, which is a face photo database. More than 13,000 facial photographs were collected from the Internet, with each face annotated with the name of the person in the shot, and 1,680 people in the sample had at least two different photos. The LFW data set is utilized to train a convolutional neural network in this paper, with some students from self-recorded films added to the data set for training.

TABLE 1: Hyperparameter settings.

Type	Hyperparameter
OptimizerA	Adam
Learning rate	0.0001
bata_1	0.99
bata_2	0.99
Epsilon	$3e-8$
Decay	$3e-8$

In addition, the data set of students' classroom behavior comes from our collection in 3 months.

**4.3. Evaluation Methods.** The quality of the teaching quality promotion strategy of college English flipped classroom depends on the accuracy of face and behavior recognition. Therefore, I use the evaluation method of classification problem to evaluate the algorithm of this article. In the classification problem, the commonly used indicators include accuracy rate, recall rate, and accuracy rate. Accuracy is the ratio of the number of samples that are predicted to be positive to all samples that are predicted to be positive. The calculation equation is as follows:



FIGURE 4: The effect of face recognition of students in classroom.

$$\text{Precision} = \frac{TP}{TP + FP}, \quad (4)$$

$$\text{Re call} = \frac{TP}{TP + FN}, \quad (5)$$

$$\text{ACC} = \frac{TP + TN}{TP + FN + FP + TN}. \quad (6)$$

**4.4. Experimental Results.** The classroom face recognition effect based on the improved VGG-16 network is shown in Figure 4. It can be seen from the figure that, except for the occlusion phenomenon of individual students, the faces of other students can be detected. In addition, it can be found that the improved VGG-16 network can also identify the information of the students in the back row, but the matching degree of the students in the back row is still flawed. The faces of individual students can be detected, but they cannot be matched. In the figure, there is a student whose face is occluded, and the algorithm has not recognized it. This is related to the lack of features provided by the student's face.

Table 2 shows the test results of the improved VGG-16 network face recognition method on the LFW data set. Compared with the classical VGG-16 network model, the accuracy of the improved VGG-16 network model in classroom face recognition is improved by 2.6%. In the study of classroom behavior recognition, student face recognition is mainly used to verify student information. For the verification of standing up students' information, the accuracy rate of face recognition reaches 96.81, fully meeting the expected requirements.

In summary, the face recognition algorithm based on the improved VGG-16 has achieved the expected effect. According to the experimental results, it can be concluded that facial features can be used as a method of confirming student identity information when performing classroom behavior recognition.

**4.5. Ablation Experiments.** In order to verify the effectiveness of softmax loss and center loss in the proposed method, an ablation experiment is set up in this section; SL stands for softmax loss, CL stands for center loss, T stands for upper

TABLE 2: Comparison result of recognition accuracy.

Method	Precision	Recall	ACC
VGG-16	93.71	84.25	84.23
VGG-19	94.22	85.26	83.25
Ours	96.81	88.01	89.65

TABLE 3: The results of the ablation experiments.

Method	Precision	Recall	ACC
T-SL-SL,B-SL-CL	92.15	81.63	83.65
T-SL-SL,B-CL-CL	91.69	82.65	84.12
T-SL-CL,B-CL-CL	90.25	84.25	85.25
T-CL-CL,B-CL-CL	92.65	84.96	81.56
T-SL-CL,B-SL-CL	96.81	88.01	89.65

TABLE 4: The results of the ablation experiments for VGG.

Method	Precision	Recall	ACC
T-VGG16,B-VGG19	91.11	82.23	84.12
T-VGG19,B-VGG16	92.25	83.69	85.12
T-VGG19,B-VGG19	91.10	83.36	84.14
T-VGG16,B-VGG16	96.81	88.01	89.65

branch, B stands for lower branch, and the results of the ablation experiment are shown in Table 3.

It can be seen from Table 3 that when the upper branch and the lower branch use softmax loss and center loss at the same time, the best performance is achieved. Therefore, this proves that the proposed method is scientific and effective.

**4.6. Ablation Study for VGG.** In this paper, a VGG ablation experiment was carried out. Considering the complexity and cost of the model, we chose VGG16 and VGG19 for the ablation experiment. The experimental results are shown in Table 4.

It can be clearly seen from Table 4 that the upper and lower branches use VGG16 to obtain the best performance, which proves that the proposed method is effective. In addition, we found that the upper branch uses VGG19 and the lower branch uses VGG16 to obtain suboptimal performance.

## 5. Conclusion

The degree of student participation in classroom learning is the main factor that restricts the effectiveness of teaching. The lack of research in this subject greatly limits the “flipped classroom” teaching model’s ability to improve college English classroom teaching quality. The degree of engagement between teachers and students, the enthusiasm of students in class, and the competence of teachers to educate are all reflected in student conduct in the classroom. Understanding and evaluating the behaviors and activities of students in the classroom are helpful in determining the state of students in the classroom, as well as improving the flipped classroom teaching technique and quality. As a result, the convolutional neural network is used to recognize student behavior in the classroom. The loss function of VGG-19 has been enhanced, the distance within the class has been lowered, the distance between the classes has been increased, and the recognition accuracy has improved. Recognizing classroom behavior aids in the development of teaching quality improvement initiatives. In addition, the experimental results show that the proposed method achieves an accuracy of 96.81% and achieves a competitive performance.

In the following research, we will focus on the processing and identification of real-time data.

## Data Availability

The data used to support the findings of this study are included within the article.

## Conflicts of Interest

The author does not have any possible conflicts of interest.

## Acknowledgments

The study was supported by the “science research project of Inner Mongolia Autonomous Region Higher Learning Institutions (Grant No. NJSY20311).”

## References

- [1] Z. Turan and B. Akdag-Cimen, “Flipped classroom in English language teaching: a systematic review,” *Computer Assisted Language Learning*, vol. 33, no. 5-6, pp. 590–606, 2020.
- [2] T. Tang, A. M. Abuhmaid, M. Olaimat, D. M. Oudat, M. Aldhaeabi, and E. Bamanger, “Efficiency of flipped classroom with online-based teaching under COVID-19,” *Interactive Learning Environments*, pp. 1–12, 2020.
- [3] R. Kay, T. Mac Donald, and M. DiGiuseppe, “A comparison of lecture-based, active, and flipped classroom teaching approaches in higher education,” *Journal of Computing in Higher Education*, vol. 31, no. 3, pp. 449–471, 2019.
- [4] P. Zheng, X. Wang, and J. Li, “Exploration and practice of curriculum ideological and political construction reform——take” information security” course as an example,” *ASP Transactions on Computers*, vol. 1, no. 1, pp. 1–5, 2021.
- [5] L. Liang, Q. Yin, and C. Shi, “Exploring proper names online and its application in English teaching in university,” *ASP Transactions on Computers*, vol. 1, no. 1, pp. 24–29, 2021.
- [6] S. S. Ismail and S. A. Abdulla, “Virtual flipped classroom: new teaching model to grant the learners knowledge and motivation,” *Journal of Technology and Science Education*, vol. 9, no. 2, pp. 168–183, 2019.
- [7] Ö. Canaran and İ. H. Mirici, “A new model of team teaching for teacher professional development: a case study of in-service English teachers,” *Egitim ve Bilim*, vol. 45, no. 201, 2020.
- [8] J. L. Jensen, E. A. Holt, J. B. Sowards, T. H. Ogden, and R. E. West, “Investigating strategies for pre-class content learning in a flipped classroom,” *Journal of Science Education and Technology*, vol. 27, no. 6, pp. 523–535, 2018.
- [9] J. Jovanovic, N. Mirriahi, D. Gašević, S. Dawson, and A. Pardo, “Predictive power of regularity of pre-class activities in a flipped classroom,” *Computers & Education*, vol. 134, pp. 156–168, 2019.
- [10] R. M. Yilmaz and O. Baydas, “An examination of undergraduates’ metacognitive strategies in pre-class asynchronous activity in a flipped classroom,” *Educational Technology Research and Development*, vol. 65, no. 6, pp. 1547–1567, 2017.
- [11] J. K. Klingner and S. Vaughn, “Promoting reading comprehension, content learning, and English acquisition through Collaborative Strategic Reading (CSR),” *The Reading Teacher*, vol. 52, no. 7, pp. 738–747, 1999.
- [12] C. McKee, “A comparison of pronouns and anaphors in Italian and English acquisition,” *Language Acquisition*, vol. 2, no. 1, pp. 21–54, 1992.
- [13] J. J. Steel, “Microbial murders crime scene investigation: an active team-based learning project that enhances student enthusiasm and comprehension of clinical microbial pathogens,” *Journal of Microbiology & Biology Education*, vol. 18, no. 2, p. 18, 2017.
- [14] K. Van Lehn, S. Ohlsson, and R. Nason, “Applications of simulated students: an exploration,” *Journal of Artificial Intelligence in Education*, vol. 5, pp. 135–135, 1994.
- [15] T. E. Dorgu, “Different teaching methods: a panacea for effective curriculum implementation in the classroom,” *International Journal of Secondary Education*, vol. 3, no. 6, p. 77, 2016.
- [16] V. A. Dinh, J. Frederick, R. Bartos, T. M. Shankel, and L. Werner, “Effects of ultrasound implementation on physical examination learning and teaching during the first year of medical education,” *Journal of Ultrasound in Medicine*, vol. 34, no. 1, pp. 43–50, 2015.
- [17] J. Chen, “Research on the promotion of Chinese vocational college students’ participation in online and offline blended teaching,” *International Journal of New Developments in Education*, vol. 2, no. 3, 2020.
- [18] P. Cooper and D. McIntyre, “Patterns of interaction between teachers’ and students’ classroom thinking, and their implications for the provision of learning opportunities,” *Teaching and Teacher Education*, vol. 10, no. 6, pp. 633–646, 1994.



- [19] N. Mercer and L. Dawes, "The study of talk between teachers and students, from the 1970s until the 2010s," *Oxford Review of Education*, vol. 40, no. 4, pp. 430–445, 2014.
- [20] J. Zhang, X. Jin, J. Sun, J. Wang, and A. K. Sangaiah, "Spatial and semantic convolutional features for robust visual object tracking," *Multimedia Tools and Applications*, vol. 79, no. 21–22, pp. 15095–15115, 2020.
- [21] R. Liu, X. Ning, W. Cai, and G. Li, "Multiscale dense cross-attention mechanism with covariance pooling for hyperspectral image scene classification," *Mobile Information Systems*, vol. 2021, Article ID 9962057, 15 pages, 2021.
- [22] W. Cai, Z. Wei, R. Liu, Y. Zhuang, Y. Wang, and X. Ning, "Remote sensing image recognition based on multi-attention residual fusion networks," *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 1, pp. 1–8, 2021.
- [23] S. Qi, X. Ning, G. Yang et al., "Review of multi-view 3D object recognition methods based on deep learning," *Displays*, vol. 69, article 102053, 2021.
- [24] P. Chonggao, "Simulation of student classroom behavior recognition based on cluster analysis and random forest algorithm," *Journal of Intelligent & Fuzzy Systems*, vol. 40, no. 2, pp. 2421–2431, 2021.
- [25] A. Vilorio, J. P. Lis-Gutiérrez, M. Gaitán-Angulo, A. R. M. Godoy, G. C. Moreno, and S. J. Kamatkar, "Methodology for the design of a student pattern recognition tool to facilitate the teaching-learning process through knowledge data discovery (big data)," in *International conference on data mining and big data*, pp. 670–679, Springer, Cham, 2018.
- [26] W. Cai, Z. Wei, Y. Song, M. Li, and X. Yang, "Residual-capsule networks with threshold convolution for segmentation of wheat plantation rows in UAV images," *Multimedia Tools and Applications*, pp. 1–17, 2021.
- [27] D. P. Campos, P. J. Abatti, F. L. Bertotti, J. A. G. Hill, and A. L. F. da Silveira, "Surface electromyography segmentation and feature extraction for ingestive behavior recognition in ruminants," *Computers and Electronics in Agriculture*, vol. 153, pp. 325–333, 2018.
- [28] J. Zhang, J. Sun, J. Wang, and X. G. Yue, "Visual object tracking based on residual network and cascaded correlation filters," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, pp. 8427–8440, 2021.
- [29] K. Simonyan and A. Zisserman, "Two-stream convolutional networks for action recognition in videos," 2014, <https://arxiv.org/abs/1406.2199>.
- [30] C. Feichtenhofer, A. Pinz, and A. Zisserman, "Convolutional two-stream network fusion for video action recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1933–1941, Las Vegas, NV, USA, 2016.
- [31] X. Ning, Y. Wang, W. Tian, L. Liu, and W. Cai, "A biomimetic covering learning method based on principle of homology continuity," *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 1, pp. 9–16, 2021.
- [32] Y. Tong, L. Yu, S. Li, J. Liu, H. Qin, and W. Li, "Polynomial fitting algorithm based on neural network," *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 1, pp. 32–39, 2021.

## Research Article

# Partial Color Photo Processing Method for Components Based on Image Enhancement Technology

Hao Wu<sup>1</sup> and Zhi Zhou<sup>2</sup>

<sup>1</sup>School of Mechatronic Engineering and Automation, Shanghai University, Shanghai 200444, China

<sup>2</sup>Samsung Electronics (China) Research and Development Center, Nanjing 210012, China

Correspondence should be addressed to Hao Wu; [yzwuhao4038@shu.edu.cn](mailto:yzwuhao4038@shu.edu.cn)

Received 30 June 2021; Revised 16 July 2021; Accepted 27 July 2021; Published 15 August 2021

Academic Editor: Shan Zhong

Copyright © 2021 Hao Wu and Zhi Zhou. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Computer vision is currently playing an increasingly important role in automatically identifying the character of the image processing technology as research hotbed in the field of smart computing, OCR, face recognition, fingerprinting, biometric recognition, and so forth. Content-based image recovery, video recovery, multimedia collection, watermarking, games, film stunts, virtual reality, e-commerce, and other apps are available all round. The color pictures of parts taken by industrial cameras depend on computer performance and the intricate environment, and in particular, on the whole resolution image display, a lot of CPU resources are needed. Some details cannot be shown completely at the same time. If the image is not sufficiently clearly visible, methods for image processing like improvement, noise reduction, and interpolation must be used to improve color photo clarity. This article, based on the OpenCV platform, uses frequency domain filters, median filters, Fourier transform, and other image improvement technologies to remove image noise in order to enhance the quality of local photos from industrial cameras' components. Finally, clear and available image information is obtained in different experimental methods, which check the application of image enhancement technology to image rebuilding. Finally, the performance of the proposed method in terms of CPBD value, definition  $Q$  value, and operation time is compared, which shows that the proposed method has obvious advantages in the above performance.

## 1. Introduction

In recent years, computer vision has been implemented in image recognition, cut, important information capture, and other areas as a consequence of the rapid development of deep learning and artificial intelligence technology. In 2014, the world-famous internet company recognized a cat from many photos through powerful computer resources which made great breakthroughs in computer technology and in the field of machine learning. With the promotion of algorithm progress, mainstream AI companies have reduced the data volume of identifying target objects from 10 million pictures to 80000, and data annotation and image processing methods play an indispensable role in the application of computer vision. In real life, it is necessary to take a lot of photos to collect a large amount of data. For industrial level data, it is difficult to ensure that all color pictures meet the

standard. Therefore, it is necessary to process the pictures twice. In digital image processing, image restoration and reconstruction is an important work. Usually, the filter algorithm is used to remove the noise in the image, and the image enhancement technology is introduced to make the image have a higher resolution and clearer image. The shooting and recognition of industrial cameras in parts are typical applications. For some defective and fuzzy images, it needs to be processed separately. If the image is not clear enough, image processing techniques such as enhancement, noise reduction, and interpolation can be used to enhance the image display resolution.

Image enhancement refers to a series of methods to improve the display effect of images, or to transform the image into a form that is easier to process by machine. From the enhanced scope, it can be divided into two types: spatial domain enhancement and frequency domain enhancement.

From the perspective of method, it can be divided into four categories: point operations (it contains image negative, contrast stretching, compression of dynamic range, gray level slicing, image subtraction, and image averaging and histogram), mask operations (smoothing, medium filtering, sharpening, and dynamic operations), transform operations (low pass filtering, high pass filtering, and high pass filtering), band pass filtering, homographic filtering, and sorting operations (false coloring and full color processing). Image enhancement can highlight the edge and important texture features, remove some noise, and suppress the display of unimportant areas, which improves the visual effect of the image to a certain extent. In fact, there are many factors that affect the quality of images, such as image degradation refers to the image quality deterioration due to the imperfection of imaging system, transmission medium, and equipment in the formation, storage, and transmission process. In order to make the enhanced image closer to the original image, it is usually necessary to analyze the reasons for image degradation, which may be caused by lens distortion, or because of motion blur, and also due to noise. Others such as the divergence caused by the optical characteristics of the imaging system, the image blur caused by noise and relative motion, and noise from circuit and photometric factors, images in space satellite, remote sensing, and astronomy will degrade due to atmospheric turbulence and relative motion between camera and object.

In the work of Kwon et al. [1], a Gaussian filtering method is proposed to enhance the image in time domain, which can improve the quality of image reconstruction. In image enhancement, edge detection and resolution enhancement are important problems in computer vision. Based on the principle of photon time stretching technology, literature [2] deduces the relationship between the contrast of the damaged image and the inherent nonlinear transfer function and proposes a visual transformation method to extract image features and enhance the display effect of the image. According to the causes of image damage, some researchers use different methods to classify the image according to the degree of blur, noise, and contrast. The typical factors are the image visual damage caused by the illumination conditions, the imaging situation of industrial cameras, and the changes of the surrounding environment. In view of the influence of fog in vehicle camera image detection, literature [3] proposed an image descriptor to distinguish whether there is fog in the photo, which solved the problem of fog image recognition and detection to a certain extent. In addition, some wavelet transform methods have also been applied to digital image processing. Literature [4] uses discrete wavelet transform and matrix singular value decomposition to reconstruct the image, which enhances the contrast of satellite image and achieves better results compared with traditional methods.

The second part of this paper summarizes the research status of image processing, the third part proposes Fourier transform and frequency domain filtering processing methods, the fourth part proposes to enhance the blurring of industrial images, and the fifth part carries out experimental verification and performance evaluation on the proposed methods.

## 2. Related Work

Recently, many machine learning technologies have been used in image processing field [5–7]. As image processing is the restoration and reconstruction of the original image, it has great application value in real life. Image enhancement can highlight the areas of interest, make the image outline clearer, and express more details. Therefore, there are many image processing [8–13] and enhancement signal processing technologies [14–16]. The traditional image enhancement technologies [17] are mainly in the form of manual extraction and construction of parameters, filtering, or transformation according to the noise and distortion degree of the image, and typical image processing technologies include GHE technology [18, 19], gray transformation, histogram, low-pass filtering, high pass filtering, and differential operation. Due to the influence of the weather environment, the rain image has low definition, accompanied by more noise and jitter delay. How to make the rain image clearer is a significant work. Literature [20] processes the image many times and obtains the single enhanced rain image through periodic filtering. For the captured image, in addition to the use of high pass filtering, channel calculation is also carried out for the processed image in each stage, so as to eliminate the rain trace and get a clearer map image. In addition, it is also necessary to enhance the display of color image. Literature [21], multiscale retinex (MSR), and guided filter (GF) are proposed for color image enhancement, and the contrast of the image is expanded by contrast lifting, so as to improve the calculation efficiency and display effect of true color image. With the rise of deep learning algorithm, some researchers try to use deep learning algorithm [22–26] to enhance the image. Literature [27] proposes a fast shot divide and glow (fsdg) network to segment and enhance the image. The effect of the image generated under weak light conditions is better than that of the artificial construction parameters. Literature [28] proposes a denoising algorithm for resisting high-frequency spatial interference to detect images with low definition. An end-to-end convolutional neural network is used to train the target image, and the target with enhanced image details is added to the loss function. The effectiveness of the method is verified on the data set, and this kind of network can detect the edge details in the image better and enhance the clarity of the image. Literature [29] proposes a convolutional neural network-based image illumination enhancement method, which uses light enhancement net to train the road scene image under low illumination conditions, converts the daytime image into weak light image, and reduces the risk of vehicle collision peak. In order to solve the problem of training data in deep learning image enhancement, an unsupervised network is proposed in literature [30], which introduces attention mechanism [31, 32] to focus on image loss sensing fusion and can enhance the image in low light level, at the same time, it can adapt to the real image enhancement display in different fields.

There are many application fields of image enhancement, such as medical treatment, industry, science and technology, and daily photo processing. Literature [33] aims at the problem that the image captured by computer in rainy day is

fuzzy and proposes a rain image enhancement method based on joint depth neural network. This method mainly uses the encoder to recognize the rain stripes in the image and uses the multilayer convolution network and the relu activation function to enhance the image. To a certain extent, it solves the low-quality problem of taking photos in rainy days at night. Literature [34] is aimed at enhancing the clarity of medical images, highlighting the local and overall characteristics of medical imaging, making the model pay attention to the region of medical interest, and using the pretraining model RESNET and efficient net model training network to enhance the image. The problem of the low contrast of chest X-ray image of pulmonary tuberculosis was solved. A more interesting work is that literature [35] studies the influence of eyelashes, glasses frame, skin, and background noise in iris recognition. In order to better detect the iris boundary, they combined convolution neural network and capsule network to propose three filtering methods for fuzzy image enhancement, It improves the accuracy of the algorithm in iris recognition task. In the above research work, the image blurring is not considered in the image processing process, and the processing process is processed according to the general image processing flow. The enhancement image processing process proposed in this paper takes into account the general image enhancement process and image blur and can enhance the clarity of the image by denoising and interpolation.

### 3. Proposed Method

The image is not clear enough, so it should again be processed, due to the limited processing capacity and distortion in the lens of the industrial camera. Filtering the frequency domain can remove the noise from a frequency angle. This is a common method of image transformation and can improve the image display resolution. This paper contains information of the function in the frequency domain with the assistance of the two-dimensional Fourier transformation, and filters enhance the image.

**3.1. Fourier Transform.** Fourier, a great mathematician, proved that the aperiodic function can be expressed by sine/or cosine multiplied by the integral of the weighted function. Any time sequence or signal of continuous measurement can be expressed as the infinite superposition of sine wave signals of different frequencies. The Fourier transform algorithm based on this principle calculates the frequency, amplitude, and phase of different sine wave signals in the signal by using the original signal measured directly. It is also similar to the image, which can transform the time domain of signal processing to the space domain, map the one-dimensional information to the two-dimensional information, and transform the sine wave to the spatial frequency wave. Therefore, the Fourier principle can be applied to the two-dimensional image processing task.

The frequency of the image is an indicator of the intensity of gray changes in the image, and it is the gradient of gray in the plane space. The spectrum map obtained by the two-dimensional Fourier transform of the image is the distribution map of the image gradient. Of course, there is no corre-

sponding relationship between the points on the spectrum map and the points on the image, even without frequency shift. In fact, the intensity of the difference between a certain point in the image and the neighboring point is the size of the gradient, that is, the size of the frequency of the point. For example, in the region of slow gray change in the image, the corresponding frequency value is very low; however, the frequency value is higher in the region where the gray level changes sharply.

For the frequency transformation degree of the image, a large gradient means that the brightness of the point is strong; otherwise, the brightness is weak. In this way, we can observe the spectrum after the Fourier transform, especially the energy distribution of the image. If there are more dark points in the spectrum, then, the actual image is softer (because the difference between each point and the neighborhood is not big, and the gradient is relatively small). On the contrary, if there are more bright points in the spectrum, then, the actual image must be sharp, and the boundary is clear and the pixels on both sides of the boundary are quite different. After shifting the frequency spectrum to the origin, we can see that the frequency distribution of the image is symmetrical with the origin as the center. In addition to clearly showing the frequency distribution of the image, shifting the frequency spectrum to the center of the circle has another advantage. It can separate interference signals with periodic regularity, such as sinusoidal interference. On a frequency spectrum with sinusoidal interference and shifting the frequency to the origin, it can be seen that in addition to the center, there is a set of symmetrically distributed bright spots with a certain point as the center, and this set is generated by the interference noise. At this time, the interference can be eliminated intuitively by placing a band stop filter at this position.

The form of two-dimensional discrete Fourier transform is as follows:

$$f(\mu, \alpha) = \sum_{p=0}^{M-1} \sum_{q=0}^{N-1} f(p, q) e^{-j2\pi(\mu p/M + \alpha q/N)}. \quad (1)$$

The corresponding inverse Fourier transform is as follows:

$$f(p, q) = \frac{1}{MN} \sum_{\mu=0}^{M-1} \sum_{\alpha=0}^{N-1} F(\mu, \alpha) e^{j2\pi(\mu p/M + \alpha q/N)}. \quad (2)$$

In the above expression,  $\mu, p \in [0, M-1]$ ,  $\alpha, q \in [0, N-1]$ . The transformed amplitude spectrum is expressed as:

$$F(\mu, \alpha) = |F(\mu, \alpha)| e^{j\varphi(\mu, \alpha)}. \quad (3)$$

Two-dimensional Fourier transform has the property of linear addition. Fourier transform is a linear system. The Fourier transform of function sum is equal to the sum of Fourier transform of each function.

$$\beta_1 f_1(x, y) + \beta_2 f_2(x, y) = \beta_1 F_1(\mu, \alpha) + \beta_2 F_2(\mu, \alpha). \quad (4)$$

According to the similarity theorem, signal compression in time domain ( $k > 1$ , faster change speed) is equivalent to signal expansion in frequency domain (widening frequency band). On the contrary, signal stretching in time domain ( $k < 1$ , slower change speed) represents signal compression in frequency domain (narrowing frequency band). The symbol is as follows:

$$f(px, qy) = \frac{1}{pq} F\left(\frac{\mu}{p} + \frac{\alpha}{q}\right). \quad (5)$$

In the Fourier transform of image, it can be divided into amplitude spectrum and phase spectrum. Amplitude spectrum represents the amount of a certain frequency component in image, and phase spectrum represents the position of frequency component in image. In order to filter in frequency domain, we usually only care about amplitude spectrum, because from amplitude spectrum, we can see that bright line reflects the gray level change of original image, this is the contour edge of the image. The Fourier transform of the image is the transformation between the spatial domain and the frequency domain.

Discrete Fourier transform is an important tool of digital signal processing, which can transform the signal information in the time domain, but it also has some disadvantages, such as large amount of calculation and long operation time. As a result, the algorithm limits its application range to some extent.

*3.2. Frequency Domain Filtering.* With Fourier transform as the basis, the image can be enhanced by frequency domain filtering. Frequency domain image enhancement is to transform the image from spatial domain to frequency domain by Fourier transform and processes the frequency components of the image accordingly, so as to realize the function of image enhancement. The spatial image enhancement based on Fourier transform is a linear system filtering:

$$f(x, y) = g(x, y) * h(x, y). \quad (6)$$

By convolution operation of Fourier transform, the above linear system can be transformed into

$$F(\mu, \alpha) = G(\mu, \alpha) * H(\mu, \alpha). \quad (7)$$

Among them,  $H(\mu, \alpha)$  represents the Fourier transform of the image  $h(x, y)$  to be enhanced; the Fourier transform  $G(\mu, \alpha)$ , which represents the unit impulse response of the spatial linear system, is called transfer function or filter function. Fourier transform is used to change the amplitude spectrum (without changing the phase spectrum) of the frequency domain image of  $f(x, y)$  so as to realize the frequency domain filtering enhancement of the image captured by the industrial camera

The principle and steps of image enhancement by frequency domain filtering are as follows:

- (1) Preprocess the image (0 fill)

$$f_0(x, y) = f(x, y) \quad 0 \leq x \leq A - 1; 0 \leq y \leq B - 1, \quad (8)$$

$$f_0(x, y) = 0 \quad A \leq x \leq P; B \leq y \leq Q. \quad (9)$$

- (2) Move the special pixel to the center

$$f_c(x, y) = f_0(x, y) * (-1)^{x+y}. \quad (10)$$

- (3) Perform Fourier transform operation on the image:

$$F(\mu, \alpha) = F * f_c(x, y). \quad (11)$$

- (4) Filtering of the original image

$$G(\mu, \alpha) = F(\mu, \alpha) * H(\mu, \alpha). \quad (12)$$

- (5) Perform the inverse Fourier transform operation again:

$$g_c(\mu, \alpha) = F_{-1} * G(\mu, \alpha). \quad (13)$$

- (6) After filtering, the special pixels are shifted back:

$$g_p(x, y) = g_c(x, y) * (-1)^{x+y}. \quad (14)$$

- (7) After filtering, extract the left quadrant

$$g(x, y) = g_p(x, y) \quad 0 \leq x \leq A - 1; 0 \leq y \leq B - 1. \quad (15)$$

Through the above 7 steps, the enhanced image after frequency domain filtering can be extracted with the help of Fourier transform. Frequency domain filtering can obtain global filtering enhancement information, obtain more ideal results and deal with more complex clutter removal multifeature enhancement, construct a filter, deliberately boost some frequency components, and depress or remove other components, so as to achieve the purpose of image enhancement, as shown in Figure 1.

*3.3. Image Enhancement Processing of Color Photos of Components Taken by Industrial Cameras.* When the local color image of components captured by industrial camera is not clear enough, the image needs to be denoised and

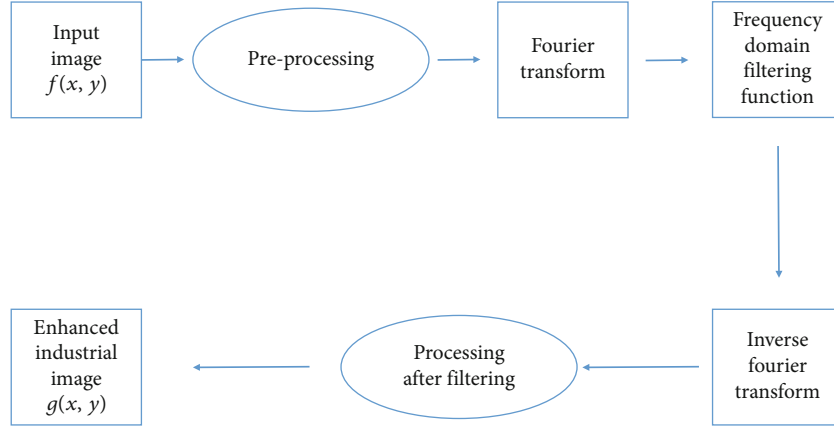


FIGURE 1: The process of components image enhancement using frequency domain filtering.

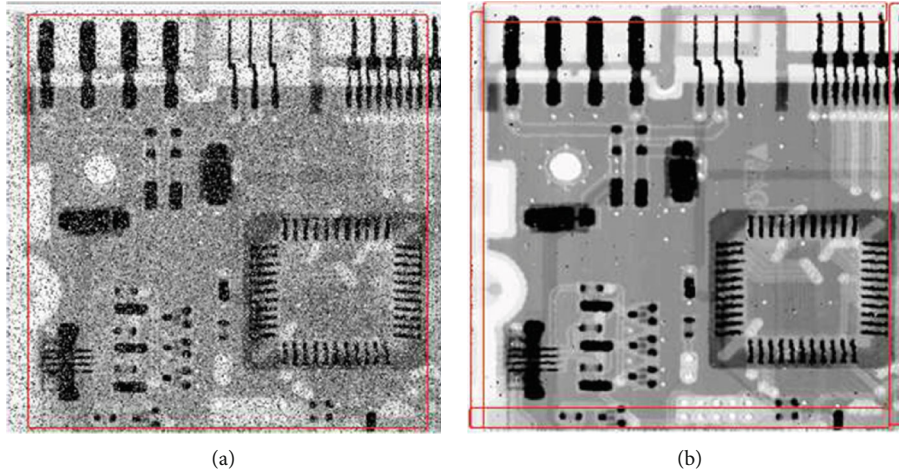


FIGURE 2: The left is the original picture of the parts taken by the industrial camera, and the right is the color image of the parts after filtering and enhancement in the frequency domain.

interpolated to enhance the resolution of the image. Because frequency domain filtering can directly process the frequency spectrum and reconstruct the display effect of the image, this paper proposes a frequency domain filtering based on Fourier transform to enhance the true color image information of components. The goal is to build a low-pass filter, which can effectively prevent the high-frequency components and allow the low-frequency components to pass smoothly, so as to filter out the noise in the high-frequency part of the frequency domain, and then the enhanced image after smooth denoising can be obtained by inverse transform.

Low pass filtering needs to calculate and intercept the frequency of the graph. First, it needs to calculate the total signal energy PE,

$$PE = \sum_{\partial}^{N-1} \sum_{\beta}^{N-1} H(\partial, \beta). \quad (16)$$

The filtering power component of the calculated image is as follows:

$$H(\partial, \beta) = |F(\partial, \beta)|^2 = R^2(\partial, \beta) + I^2(\partial, \beta). \quad (17)$$

In  $H(x, y)$  with low definition image whose coordinates are  $(x, y)$ , the image filtered in frequency domain is as follows:

$$T(x, y) = F(x, y) * H(x, y). \quad (18)$$

## 4. Experiment and Results

**4.1. Experimental Realization.** Based on the principle of frequency domain filtering for image enhancement, we randomly selected 100 true color images of components from the blurred images captured by industrial cameras and carried out experiments on the OpenCV platform to obtain the images after low-pass filtering. In Figure 1, the image on the left is taken by an industrial camera due to the influence of noise. From the image, the global information becomes very fuzzy and cannot reflect the real information of the image. Therefore, image enhancement is needed. In the experiment, the texture information of the image can be restored and reconstructed by using Fourier transform and frequency domain filtering. The reconstructed image is shown in the right figure, after filtering and enhancement,



FIGURE 3: Performance comparison under the evaluation index of image A.

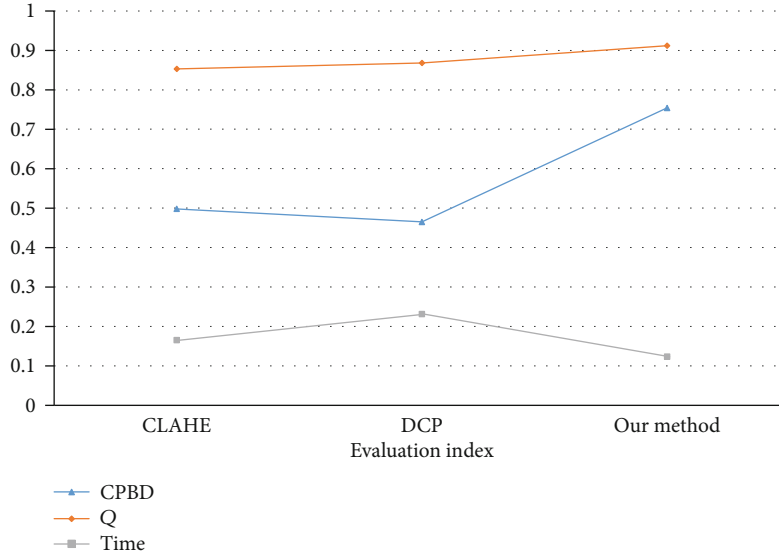


FIGURE 4: Performance comparison under the evaluation index of image B.

the global image information can be clearer, and the edge information can also be highlighted, as shown in Figure 2.

**4.2. Performance Evaluation.** At present, the evaluation methods for blurred image enhancement technology, because of the essence of images, restore some detail information or edge information for blurred images or low-resolution images. Therefore, this paper chooses Cumulative Fuzzy Detection Probability Measurement Index (CPBD) and Clarity Q Index for comprehensive evaluation.

CPBD index is a contrast measurement index based on cumulative blur detection probability without any reference to other standard images, which is the real response to blurred images after processing under visual perception. The experimental results show that there is a good correlation between this index and human consciousness evaluation.

JNB is defined as the minimum amount of perceived blur around the edge when the contrast is higher than the perceptible difference. For a given contrast, the probability of edge blur detection  $P_{\text{Blur}}$  is defined as:

$$P_{\text{Blur}} = 1 - e^{-|w(e_i)/w_{\text{JNB}}(e_i)|^\beta}, \quad (19)$$

where  $e_i$  represents the edge pixels of the blurred image; the parameter is empirically set to 3.45.  $w(e_i)$  is the measured width of the edge pixels of the blurred image;  $w_{\text{JNB}}(e_i)$  is the directly perceivable blurred edge width of the edge pixel  $e_i$  of the blurred image, and its size is related to the local contrast around the edge. The calculation formula is defined.

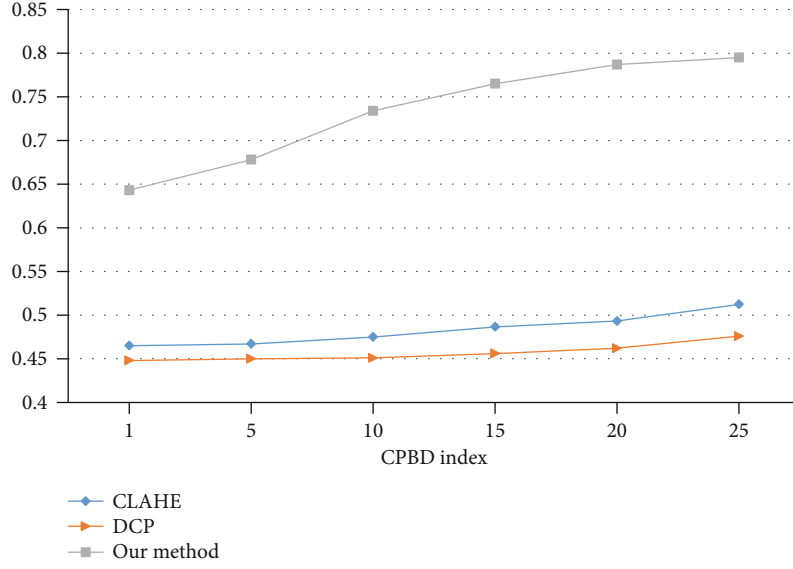


FIGURE 5: Performance comparison under evaluation index CPBD.

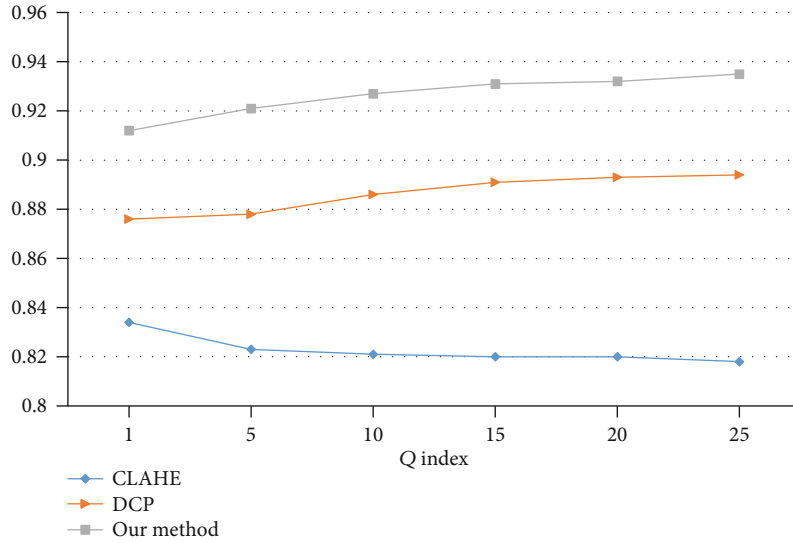


FIGURE 6: Performance comparison under evaluation index Q.

$$W_{\text{JNB}} = \begin{cases} 5, & C \leq 50, \\ 3, & C > 50, \end{cases} \quad (20)$$

where  $C$  is the local contrast of the blurred image, which is the difference between the maximum and minimum values in the edge area of the image.

Formal model description of CPBD:

$$\text{CPBD} = \sum_{P_{\text{Blur}}=0}^{P_{\text{JNB}}} P(P_{\text{Blur}}). \quad (21)$$

Another indicator is clarity, which indicates the recovery of blurred images on detail pixels, generally showing the size of resolution and edge arts and sciences. This information

can explain the image nodule recovery ability under enhancement technology from another side, described as:

$$Q = 1 - \frac{1}{3MN} \sum_{C \in (R,G,B)} \sum_{x=1}^M \sum_{y=1}^N \min[q_c(x,y), 1 - q_c(x,y)], \quad (22)$$

where  $Q$  represents the definition measure value of the image, and  $M$  and  $N$  represent the pixels of the image, respectively. It is described by the following formula:

$$q_c(x,y) = \sin \left[ \frac{\pi}{2} \times \left( 1 - \frac{f_c(x,y)}{f_{\text{max}}} \right) \right], \quad (23)$$



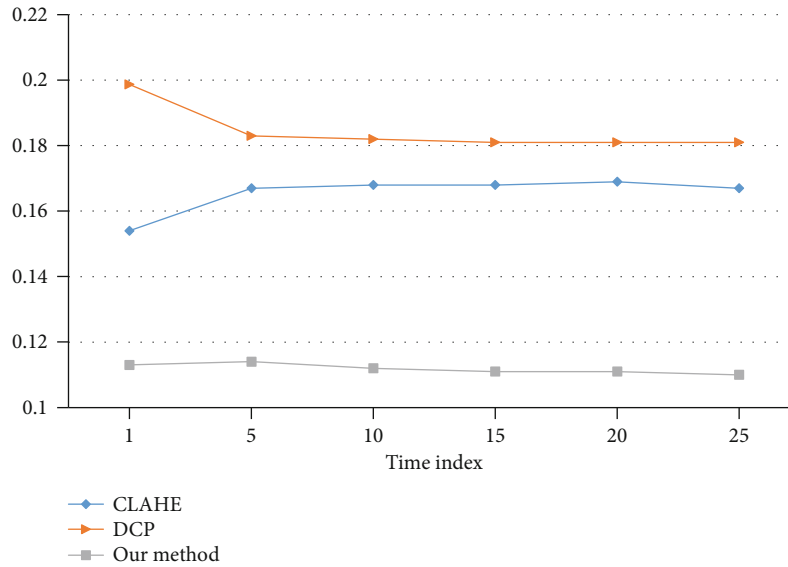


FIGURE 7: Performance comparison under evaluation index time.

where  $f_c(x, y)$  is the original image pixel value at the image  $C$  channel  $(x, y)$ ;  $f_{\max}$  represents the maximum pixel value of the original image.

**4.3. Experimental Performance Comparison.** In order to reflect the implementation effect of this algorithm, a variety of blurred images are used as experimental objects, and the performance under different images is compared under three performance indexes: CPBD value, definition  $Q$  value, and operation time. Figures 3 and 4 show the performance comparison under different blurred images.

Figures 3 and 4 show that the CPBD value of this algorithm is obviously higher than the other two algorithms, which shows that the performance of this algorithm is the best in contrast improvement; the clarity  $Q$  value of this algorithm is also the largest, which shows that the performance of this algorithm is the best in detail information recovery. In terms of the operation time, the algorithm in this paper takes the shortest time, which is much faster than the DCP algorithm.

Compared with the above performance, the performance advantage of this paper is very obvious. Therefore, this paper uses multiple groups of blurred images to analyze the performance and then takes the average value. The effect is shown in Figures 5–7.

Figures 5–7 show that the algorithm proposed in this paper is compared with CLAHE and DCP algorithms. The algorithm in this paper is obviously higher than the other two algorithms in three performance indexes, which shows that the algorithm has the best performance in contrast improvement; CLAHE is equivalent to DCP in CPBD performance. CLAHE has the worst performance under  $Q$  index, while DCP has the worst time performance under Time Index.

## 5. Conclusion and Prospect

Aiming at the problem of unclear parts pictures taken by industrial cameras, this paper proposes Fourier transform

and frequency domain filtering as algorithms to enhance the image to perform secondary processing on the photos, so that the filtered image is highlighted. The quality of the image is improved. Compared with the original image, the effect of visual effects on the quality of the final image is improved. This method can show the texture information of the target image more clearly. Due to the limitation of computing resources, we did not use deep learning to automatically process image information. In the future, the images we need to process are more complex, which may be image reconstruction in stereo scenes or image shadow noise removal under fast motion. Try to use convolutional neural network and semantic segmentation algorithms to further process the image, let the algorithm automatically acquire and learn the reconstruction method of the image, and enhance the contrast of the image, which can be widely promoted in practical applications.

## Data Availability

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Conflicts of Interest

The authors declared that they have no conflicts of interest regarding this work.

## References

- [1] S. Kwon, H. Lee, and S. Lee, "Image enhancement with Gaussian filtering in time-domain microwave imaging system for breast cancer detection," *Electronics Letters*, vol. 52, no. 5, pp. 342–344, 2016.
- [2] M. Suthar, M. Asghari, and B. Jalali, "Feature enhancement in visually impaired images," *IEEE Access*, vol. 6, pp. 1407–1415, 2017.

- [3] M. Pavlic, H. Belzner, G. Rigoll, and S. Ilić, "Image based fog detection in vehicles," in *2012 IEEE Intelligent Vehicles Symposium*, Madrid, Spain, 2012.
- [4] H. Demirel, C. Ozcinar, and G. Anbarjafari, "Satellite image contrast enhancement using discrete wavelet transform and singular value decomposition," *IEEE Geoscience & Remote Sensing Letters*, vol. 7, no. 2, pp. 333–337, 2010.
- [5] M. Gao, W. Cai, and R. Liu, "AGTH-Net: attention-based graph convolution-guided third-order hourglass network for sports video classification," *Journal of Healthcare Engineering*, vol. 2021, Article ID 8517161, 10 pages, 2021.
- [6] Y. Jiang, X. Gu, D. Wu et al., "A novel negative-transfer-resistant fuzzy clustering model with a shared cross-domain transfer latent space and its application to brain CT image segmentation," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 18, no. 1, pp. 40–52, 2021.
- [7] Z. Zhongxiang, F. Chenghua, Z. Liang, and M. Kong, "K algorithm for microstrip peocessor design," *ASP Transactions on Neural Information Computing*, vol. 1, no. 1, pp. 15–21, 2021.
- [8] W. Cai and Z. Wei, "Remote sensing image classification based on a cross-attention mechanism and graph convolution," *IEEE Geoscience and Remote Sensing Letters*, 2020, In press.
- [9] J. Zhang, Y. Liu, H. Liu, and J. Wang, "Learning local–global multiple correlation filters for robust visual tracking with Kalman filter redetection," *Sensors*, vol. 21, no. 4, p. 1129, 2021.
- [10] X. Zhang, Y. Yang, Z. Li, X. Ning, Y. Qin, and W. Cai, "An improved encoder-decoder network based on strip pool method applied to segmentation of farmland vacancy field," *Entropy*, vol. 23, no. 4, article e23040435, p. 435, 2021.
- [11] W. Cai, Z. Wei, R. Liu, Y. Zhuang, Y. Wang, and X. Ning, "Remote sensing image recognition based on multi-attention residual fusion networks," *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 1, pp. 1–8, 2021.
- [12] J. Zhang, X. Jin, J. Sun, J. Wang, and K. Li, "Dual model learning combined with multiple feature selection for accurate visual tracking," *IEEE Access*, vol. 7, pp. 43956–43969, 2019.
- [13] J. Zhang, J. Sun, J. Wang, and X. G. Yue, "Visual object tracking based on residual network and cascaded correlation filters," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–14, 2020, In press.
- [14] Y. Jiang, Y. Zhang, C. Lin, D. Wu, and C. T. Lin, "EEG-based driver drowsiness estimation using an online multi-view and transfer TSK fuzzy system," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 3, pp. 1752–1764, 2021.
- [15] S. Qi, X. Ning, G. Yang et al., "Review of multi-view 3D object recognition methods based on deep learning," *Displays*, vol. 69, p. 102053, 2021.
- [16] M. Zheng, "PCNN for power distribution network," *ASP Transactions on Neural Information Computing*, vol. 1, no. 1, pp. 18–21, 2021.
- [17] L. Zhang, X. Wang, X. Dong, L. Sun, W. Cai, and X. Ning, "Finger vein image enhancement based on guided tri-Gaussian filters," *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 1, pp. 17–23, 2021.
- [18] T. Kim and H. S. Yang, "A multidimensional histogram equalization by fitting an isotropic Gaussian mixture to a uniform distribution," in *IEEE International Conference on Image Processing*, Atlanta, GA, USA, 2007.
- [19] L. Li, "DSP for condition monitoring equipment in real-time," *ASP Transactions on Neural Information Computing*, vol. 1, no. 1, pp. 22–25, 2021.
- [20] Z. Shi, Y. Li, M. Zhao, Y. Feng, and L. He, "Multi-stage filtering for single rainy image enhancement," *IET Image Processing*, vol. 12, no. 10, pp. 1866–1872, 2018.
- [21] H. Sadia, F. Azeem, H. Ullah, Z. Mahmood, S. Khattak, and G. Z. Khan, "Color image enhancement using multiscale retinex with guided filter," in *2018 International Conference on Frontiers of Information Technology (FIT)*, Islamabad, Pakistan, 2018.
- [22] Y. Gu, A. Chen, X. Zhang, C. Fan, K. Li, and J. Shen, "Deep learning based cell classification in imaging flow cytometer," *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 2, pp. 18–27, 2021.
- [23] Z. Huang, P. Zhang, R. Liu, and D. Li, "Immature apple detection method based on improved Yolov3," *ASP Transactions on Internet of Things*, vol. 1, no. 1, pp. 9–13, 2021.
- [24] X. Ning, K. Gong, W. Li, and L. Zhang, "JWSAA: joint weak saliency and attention aware for person re-identification," *Neurocomputing*, vol. 453, pp. 801–811, 2020.
- [25] Y. Tong, L. Yu, S. Li, J. Liu, H. Qin, and W. Li, "Polynomial fitting algorithm based on neural network," *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 1, pp. 32–39, 2021.
- [26] C. Yan, G. Pang, X. Bai et al., "Beyond triplet loss: person re-identification with fine-grained difference-aware pairwise loss," *IEEE Transactions on Multimedia*, 2021, In press.
- [27] R. Khan, Q. Liu, and Y. Yang, "A deep hybrid few shot divide and glow method for ill-light image enhancement," *IEEE Access*, vol. 9, pp. 17767–17778, 2021.
- [28] A. S. Shamsabadi, C. Oh, and A. Cavallaro, "EdgeFool: an adversarial image enhancement filter," 2019, <https://arxiv.org/abs/1910.12227>.
- [29] G. Li, Y. Yang, X. Qu, D. Cao, and K. Li, "A deep learning based image enhancement approach for autonomous driving at night," *Knowledge-Based Systems*, vol. 213, article 106617, 2020.
- [30] Y. Jiang, X. Gong, D. Liu et al., "EnlightenGAN: deep light enhancement without paired supervision," *IEEE Transactions on Image Processing*, vol. 30, pp. 2340–2349, 2019.
- [31] W. Cai, B. Liu, Z. Wei, M. Li, and J. Kan, "TARDB-net: triple-attention guided residual dense and BiLSTM networks for hyperspectral image classification," *Multimedia Tools and Applications*, vol. 80, no. 7, pp. 11291–11312, 2021.
- [32] R. Liu, X. Ning, W. Cai, and G. Li, "Multiscale dense cross-attention mechanism with covariance pooling for hyperspectral image scene classification," *Mobile Information Systems*, vol. 2021, Article ID 9962057, 15 pages, 2021.
- [33] Z. Shi, Y. Feng, M. Zhao, and L. He, "A joint deep neural networks-based method for single nighttime rainy image enhancement," *Neural Computing and Applications*, vol. 32, no. 7, pp. 1913–1926, 2020.
- [34] K. Munadi, K. Muchtar, N. Maulina, and B. Pradhan, "Image enhancement for tuberculosis detection using deep learning," *IEEE Access*, vol. 8, pp. 217897–217907, 2020.
- [35] M. Liu, Z. Zhou, P. Shang, and D. Xu, "Fuzzified image enhancement for deep learning in iris recognition," *IEEE Transactions on Fuzzy Systems*, vol. 28, no. 1, pp. 92–99, 2020.

## Research Article

# Cloud Education Chain and Education Quality Evaluation Based on Hybrid Quantum Neural Network Algorithm

Hong-Xia Liu <sup>1</sup>, Yong-Heng Zhang,<sup>1</sup> and Sang-Bing Tsai <sup>2</sup>

<sup>1</sup>School of Information Engineering, Yulin University, Shaanxi 719000, China

<sup>2</sup>Regional Green Economy Development Research Center, School of Business, WUYI University, China

Correspondence should be addressed to Hong-Xia Liu; 17719630982@163.com and Sang-Bing Tsai; sangbing@hotmail.com

Received 8 June 2021; Revised 7 July 2021; Accepted 22 July 2021; Published 15 August 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Hong-Xia Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper proposes the functional model and application service implementation process of the education cloud platform application service architecture. The entire cloud application service architecture mainly includes four parts: cloud service management, cloud application service rapid creation and deployment, dynamic process configuration, and unified identity authentication. Based on the basic theory of workflow, the process status and business services of cloud application services are discussed. The BP neural network weight optimization model based on the improved quantum evolution method is studied, and a method that combines the improved quantum evolution algorithm (IQEA) and the BP algorithm to complete the back propagation neural network training is proposed, that is, the IQEA-BP algorithm. Firstly, the traditional quantum evolution algorithm is improved, and then, the improved quantum evolution algorithm is used to optimize the network weights as a whole to overcome the shortcomings of the BP algorithm that is easy to fall into the local optimum; then, we use the BP algorithm to find the better weight as the initial value to improve the training and prediction accuracy of the network. In order to enrich the school education quality evaluation system, this article adds soft indicators that can reflect school education performance on the basis of the existing “National Education Inspection Team” indicators and uses analytical methods to prove the effectiveness and feasibility of the new evaluation indicators. The X1-X10 index data is selected as the evaluation index of the school education quality evaluation system in this paper. Testing the performance of the BP neural network, the accuracy rate of the school education quality evaluation is 93.3%, the average absolute error is 0.067, and the accuracy and recall rate of the test set grade gradient of 0, 1, 2, 3, 5, 6, and 8 are all 93%, indicating that the IQEA-BP neural network algorithm has a good effect on the evaluation of school education quality.

## 1. Introduction

The classroom teaching process and the supervision of the teaching quality in the classroom are an effective method to ensure the quality of classroom teaching [1]. The direct performance of teachers' classroom teaching effect is the students' mastery and application of knowledge; in order to ensure the quality of teachers' classroom teaching, it is necessary to strengthen the control of teaching quality [2]. Teaching evaluation is one of the most commonly used methods in teaching supervision activities. It is to objectively inspect and judge both the teacher's teaching process in the classroom

and the student's learning quality [3]. Through the supervision and assessment of students' learning quality, the results of the implementation can be fed back to teachers, so that they can improve or strengthen certain aspects of teaching, so as to better and more timely ensure that teachers complete teaching tasks within the corresponding teaching timeliness. Nowadays, in order to continuously improve the learning ability of students and the teaching ability of teachers, a comprehensive and systematic teaching quality monitoring system should be established and improved, which can play a positive role in cultivating talents who adapt to the times and have innovative spirit and practical ability [4].

The information network center opens the port, breaking through the geographical limitations of the original paper-based evaluation. Users can log in to the evaluation system in various forms at any time to evaluate the objects or content that need to be evaluated. This makes teaching evaluation more flexible. In today's society where informatization continues to develop, information technology can be used to establish a more complete teaching evaluation system [5]. Through this system, teachers' teaching conditions can be monitored in real time, and teachers can understand how they are teaching in a shorter period of time. The real-time situation in the process can help teachers adjust their teaching plans in a shorter time and better meet the teaching needs of students. The popularization of information technology in every industry is an inevitable trend in the development of contemporary information society. The teaching evaluation system established by information technology not only is a manifestation of social and technological progress but also can improve the quality of teaching better and faster [6].

This paper analyzes the existing cloud application service architecture and model of cloud application services, puts forward the architecture model of education cloud platform application services, and elaborates the application service implementation process. This paper introduces the method of initial weight optimization of a neural network, analyzes and discusses the advantages and disadvantages of various methods, and proposes a combination of improved quantum evolution algorithm (IQEA) and BP algorithm to complete the back propagation neural network. The algorithm first improves the traditional quantum evolution algorithm and then uses the improved quantum evolution algorithm to optimize the network weights as a whole, to overcome the shortcomings of the BP algorithm of easily falling into the local optimum. The BP algorithm is further optimized to improve the training and prediction accuracy of the network. This article uses analytical methods to classify the educational quality of some schools in a certain city. Through the evaluation and analysis of the old and new evaluation indexes, the appropriate evaluation indexes are determined. We calculate the evaluation scores of the comprehensive principal components and obtain the evaluation results of the education quality of each school. The comparison and analysis of the evaluation results of the new and old evaluation indicators show that the accuracy of the new evaluation index is higher than that of the old evaluation index, and the average absolute error is lower than that of the old evaluation index, which proves the feasibility and effectiveness of the new evaluation index. This paper selects 11 indicators of compulsory education balance index X1-X10 and the new school admission score X11 as the evaluation index of the school education quality evaluation system, which provides good data support for school education quality evaluation. We use the BP neural network evaluation model to comprehensively evaluate the quality of school education and explore the important and difficult points of the BP neural network evaluation model based on the compulsory education balance index, including the network structure design of the BP neural network and the determination of the number of neurons in the hidden layer.

## 2. Related Work

Relevant scholars merged quantum computing with traditional neural computing and proposed a quantum competitive learning algorithm capable of pattern classification and associative memory, expounding the dynamics of quantum neural network and its application in information security [7]. Researchers use universal quantum logic gates as calculation basis functions to propose a quantum neural computing model [8]. The simulation results show that this model is better than the traditional neural network model. Relevant scholars analyzed the motivation and form of the evolution of the artificial neural network vector subneural network, the advantages of the quantum neural network, and possible implementation methods and expounded the preliminary application of the quantum neural network in pattern recognition, entanglement calculation, function approximation, and so on [9]. Related scholars introduced the quantum neural network with multilayer activation function into the multisensor information fusion and proposed a multisensor information fusion integrated circuit fault diagnosis algorithm based on the quantum neural network [10].

Related scholars have studied the weight learning problem of a quantum neural network model. Researchers proposed a quantum self-organizing feature mapping network model and clustering algorithm, studied the application of the quantum neural network in pattern recognition and multimode high probability Grover algorithm, and achieved good application results [11]. Related scholars proposed a quantum BP neural network model and its learning algorithm based on the evolution of a universal quantum gate and studied the application of quantum M-P and perceptron network models as well as quantum Hopfield in image processing [12]. Based on the in-depth study of quantum revolving gates and control NOT gates, the researchers proposed a quantum gate circuit-based quantum control NOT gate neural network model and algorithm [13]. At the same time, they proposed a quantum genetic algorithm and applied it to fuzzy neural control. Based on the quantum self-organizing neural network model, the original parallel quantum neural network has been studied more deeply, and good application results have been achieved. Related scholars apply a quantum gate neural network to the field of image compression and optimize the network weights through genetic algorithm, which improves the overall performance of the network [14].

Cloud computing originated in the United States and later spread to many countries including Singapore, Japan, the United Kingdom, and India. Cloud computing first appeared as a commercial service, and then, new cloud computing technology began to be continuously developed under the promotion of Google, IBM, Amazon, and other world companies, enriching the functions of cloud computing technology and opening cooperation projects with universities. The platform uses cloud computing to serve school management and teaching development, save management costs, broaden interschool and intraschool interaction channels between teachers and students, and expand the use of resources [15]. Cloud computing was first applied to large

Internet companies and e-commerce companies. Since then, Tsinghua University and Google have cooperated to launch a cloud computing technology promotion plan and strengthen the connection between education and cloud computing technology by building a cloud computing-based teaching network platform [16].

Many companies such as China Telecom and Huawei have cooperated with universities to develop “Education Cloud” platform projects to expand the coverage of high-quality educational resources and promote educational equity. In addition, the state has launched cloud education programs such as the coconstruction and sharing of classroom teaching resources, electronic schoolbags, and dual classroom teaching applications in the pilot education cloud. Yunwei Technology Co., Ltd. has developed the world’s first cloud computing application platform dedicated to the field of education—the Yun Education Network, which provides one-stop teaching organization and management services [17]. At present, many colleges and universities have built their own educational information service platforms to provide services for teaching. In addition, many educational institutions and Internet companies have begun to develop and promote their own teaching platforms. However, there are still few complete education platforms based on cloud storage, many basic functions of the system are not yet complete, and various platforms cannot communicate with each other.

Relevant scholars pointed out that blindly adopting qualitative or quantitative methods for teaching evaluation is not comprehensive [18]. Only some general data can be obtained. In order to obtain more comprehensive data, a fuzzy comprehensive evaluation method combining qualitative and quantitative methods can be used. The analytic hierarchy process can make a better and more comprehensive evaluation of teaching quality. The use of effective evaluation methods is a prerequisite to ensure that the evaluation results are objective, fair, and accurate. Some schools use quantitative teaching evaluation forms and trust the results of a large amount of data [19]. The use of quantitative teaching evaluation forms only simplifies the content of the evaluation at the beginning, but the evaluation process needs to consume more students. It takes a lot of time for calculating data when summing up the evaluation results. Time is wasted and accurate data cannot be obtained. Some schools use the form of “listening to lectures” to evaluate teaching, but this way will affect the teacher’s teaching plan [20]. The teacher’s teaching plan has been set at the beginning of the semester. During the teaching process, it is constantly based on the situation of the students. However, the school adopts this kind of evaluation method. If teachers want to get a better evaluation, they need to take out the best state of the usual class. It is inevitable to conduct a drill with the students, and this kind of evaluation method will not reach it at all. At present, domestic universities and colleges mainly adopt the form of questionnaires supported by large amounts of data and the form of online evaluation using network information as support. Both evaluation forms have been baptized by time and are always under the continuous development of time and society. The one with the lower utilization rate will be eliminated, and a

more feasible way will be adopted to evaluate the teaching of teachers.

### 3. Education Cloud Platform Application Service Architecture Modeling

*3.1. Analysis of Application Service Architecture of Education Cloud Platform.* The education cloud platform application service refers to the education application service resources running on the cloud platform. The deployment of education application resources on the cloud platform is mainly to reduce the cost of education resources, provide resource sharing rates, and solve the problems of uneven distribution of education resources. For the education cloud platform application service as the application of software as a service in the education field, its architecture model follows the basic architecture of software as a service, and as a special case of active service, its architecture model also retains the characteristics of the active service model. From the above two points, it can be concluded that the architecture of cloud application services is as shown in Figure 1.

The basic model of cloud application services includes three roles and five basic operations. The three roles are service registry, service user, and service provider [21–23]. The five basic operations are publishing, authorization, search, binding, and execution. This architecture corresponds to the payment model of cloud application services “on-demand leasing” and provides a solution for the integrated utilization of educational resources. In this architecture model, the existing education application system can be provided as a service to schools or educational institutions. They order services according to their own needs, and after obtaining service authorization, they can access software services through a browser or client. Based on the analysis of the existing application service software architecture foundation, combined with the special needs of education cloud platform application services, a complete cloud application service architecture should have the following characteristics:

- (1) *Multitenancy*: cloud application services are software-as-a-service based on cloud platforms that are released by service providers and hosted by the service registry. This mode of operation requires cloud application services to support multitenancy in order to maximize the interests of the three parties.
- (2) *Configurability*: according to changes in tenants’ demand for cloud application services, tenants can configure the functions in the service to meet different needs.
- (3) *Openness*: the application service resources on the education cloud platform should be open and shareable, so that the maximum use of educational resources and the sharing of educational resources can be realized to a certain extent.
- (4) *Ease of use*: based on the characteristics of openness, to realize resource sharing between different application systems, the convenience of switching between different application systems must be considered.

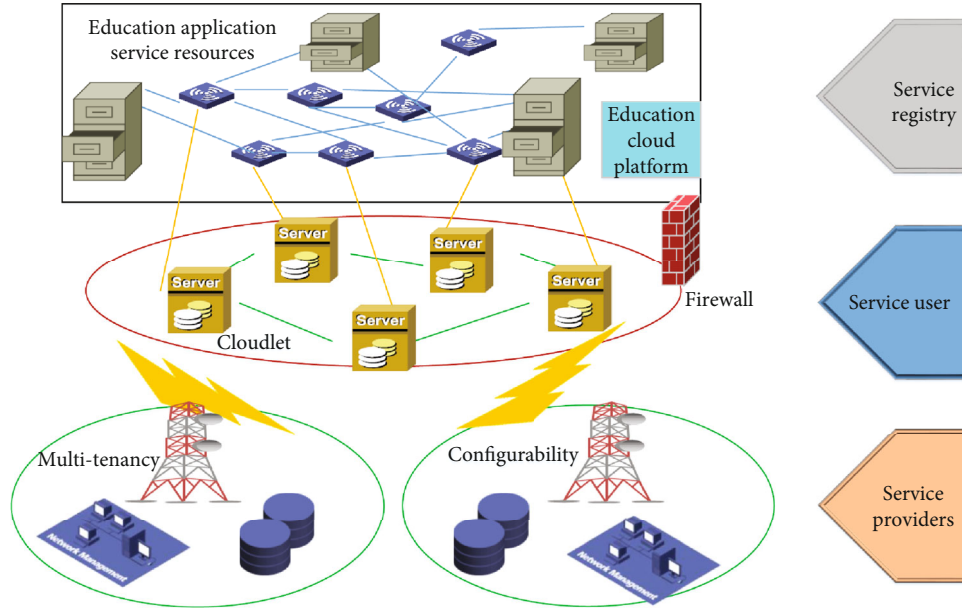


FIGURE 1: Cloud application service architecture.

### 3.2. Cloud Application Service Architecture Function Model.

To provide application services based on the education cloud platform, we must first build a cloud service management platform with the function of a service registry. Based on the idea of active service, educational application software is provided as a service for tenants to choose, and a platform interface is provided for tenants to order application services; it also provides a management platform for application software released by service providers. Secondly, application services based on the cloud platform have the characteristics of multitenancy, so it must be able to configure the service and realize the convenient deployment of the service. Third, the ultimate goal of this topic is to realize the sharing of educational resources, so a convenient and effective resource utilization plan should be proposed. According to the above considerations, an architecture model based on the education cloud platform application service is proposed, as shown in Figure 2.

As can be seen from the architecture model diagram, the cloud application service architecture has four main parts: unified authentication platform, application service configuration, service publishing, and service management center [24–26].

- (1) *Unified certification platform*: this part is responsible for the identity authentication of all application services. Users only need to have an application service account to access the shared resources of other trusted application services without additional account registration.
- (2) *Application service configuration*: after the user subscribes to the service, according to his own individual needs, the initial application service related process is set to meet the user's dynamic requirements.

- (3) *Application service release*: as the administrator of cloud application services, when users purchase corresponding application services on demand, these services must be published on the server for users to use. Service publishing provides relatively quick service deployment functions and education for service managers.

- (4) *Application service management center*: this part is similar to the service registration center of active service, and its function is to manage service information and equipment information, service order management, service maintenance, and tenant management. Users can learn about the types of education application services and service functions that exist on the cloud platform through the service management center, such as office automation management system, library management system, and student attendance management system, and then complete the service ordering process.

For service managers, it provides service providers with the software, hardware operation platform, and network infrastructure required for service operation and collects the rental fees paid by service users when they rent the service. Although in the short term, all the funds invested will not be recovered, in the long run, the benefits brought by cloud service applications are considerable (see Figure 2).

### 3.3. Cloud Application Service Implementation Process.

According to the above analysis of cloud application service architecture and the proposed cloud application service architecture function model, it can be understood that cloud application services are service software deployed on the cloud platform. These application software are managed

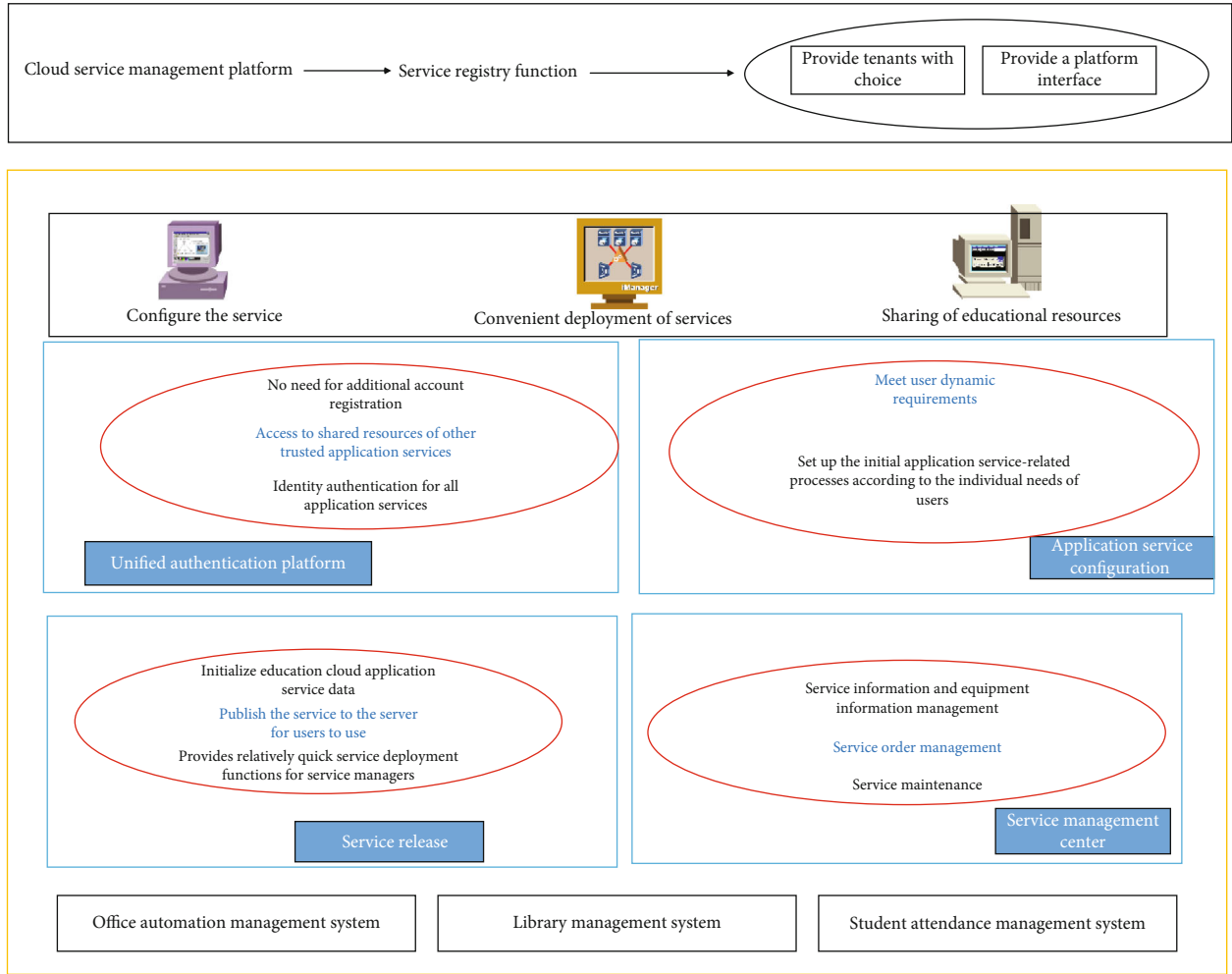


FIGURE 2: Education cloud platform application service architecture functional architecture.

and released by the service center in a unified manner, users can configure the functional flow of the service, and mutually trusted services can be free to log in through a unified identity authentication platform to realize cloud application service resource sharing. We elaborated based on the entire implementation phase of application services, and the implementation process based on the education cloud platform application service architecture is shown in Figure 3.

The overall implementation process of the application service architecture is as follows: First, the application service information is released to the service management center, and then, the user orders the cloud application service. The service management center reviews the cloud application service order and deploys the service on the cloud platform. The user initiates a request to the application service through the Internet, and the unified identity authentication part verifies the validity of the user's identity. After the verification is completed, the user can share other application service resources through the identity authentication platform without login, thereby realizing the education cloud application service software resources shared.

#### 4. BP Neural Network Weight Optimization Hybrid Model Based on Improved Quantum Evolution Method

4.1. *Improved Quantum Evolution Algorithm.* IQEA uses the quantum revolving door update strategy to complete the update operation of the population. The value of the rotation angle  $\theta_i$  is

$$\theta_i = \Delta\lambda_{i-1} \cdot S(\alpha_{i-1}, \beta_{i-1}). \quad (1)$$

Among them,  $S(\alpha_{i-1}, \beta_{i-1})$  and  $\Delta\lambda_{i-1}$  are the rotation direction and angular step of the rotation angle, respectively. Delta is a coefficient related to the convergence speed of the algorithm, and the value must be reasonable. Combined with the idea of dynamically adjusting the rotation angle of the quantum gate, we propose a specific implementation form of delta:

$$\text{delta} = 0.02\pi - \frac{(n-1) \cdot k}{\text{MAXGEN}-1}. \quad (2)$$

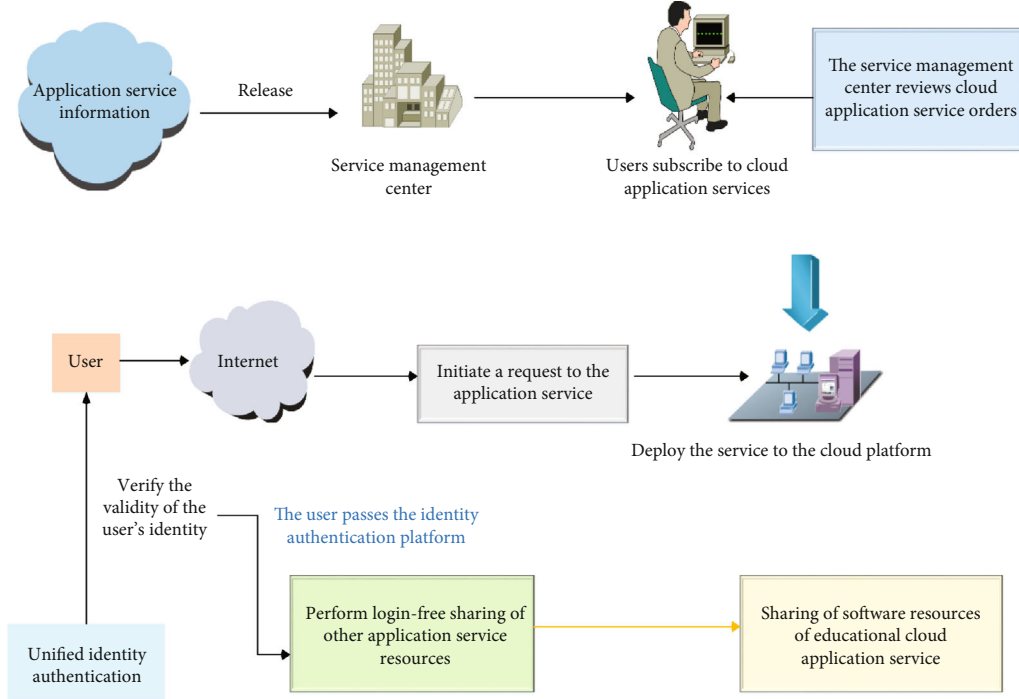


FIGURE 3: The implementation process of education cloud platform application services.

Among them,  $n$  is the current evolutionary algebra, MAXGEN is the terminating algebra, and  $k$  is a constant between  $[0,1]$ . At the beginning of the algorithm operation, the search grid is larger, which increases the convergence speed of the algorithm. At the end of the algorithm operation, the search grid is smaller, so as to achieve precise search and to help find the optimal solution.

The search capability of EA has been greatly improved through crossover operations. The quantum crossover operation introduced by IQEA makes full use of the interference of quantum states to make full use of all chromosomal information to generate more new patterns, greatly improving the search ability of the algorithm, and can effectively solve optimization problems with little correlation between genes. This paper proposes a specific implementation process of quantum crossover: randomly sort all individuals in the population; cyclically shift the  $i$ -th position of all individuals after sorting by  $i-1$  times to obtain a new population after the crossover operation.

IQEA uses quantum mutation operations to prevent the algorithm from falling into a local optimal solution. Due to the entanglement nature of quantum states, single-point mutation can effectively prevent premature maturity during quantum mutation. Therefore, this article adopts a single-point mutation quantum mutation operation.

**4.2. Artificial Neural Network and BP Algorithm.** A neural network system is a complex network system that is widely interconnected by a large number of simple processing units (neurons), which has certain learning, memory, and calculation capabilities. Neural networks have some notable features, such as nonlinear mapping capabilities, no need

for precise mathematical models, and easy implementation of software and hardware.

A multilayer perceptron (MLP) network is one of the common neural network models. This type of neural network can have multiple hidden layers, the basis function  $u(\cdot)$  takes a linear function, and the activation function  $f(\cdot)$  can take many forms. When the activation functions of all computing nodes take the hard limit function, it is called a multilayer discrete perceptron: when the hidden layer nodes take the sigmoidal function, it is a well-known BP neural network.

A BP neural network has a high degree of nonlinear mapping ability. It has been proven that a 3-layer BP network can approximate a continuous function with arbitrary precision. The activation function of hidden nodes of the BP network is a sigmoidal function, so the BP network is also called an activation function as the sigmoidal multilayer perceptron. The activation function of the output node varies depending on the purpose. When the network is used for classification, the output node activation function generally uses a sigmoidal function or a hard limit function; if it is used for function approximation, the output node activation function generally uses a linear function.

Suppose that the first hidden layer has  $n$  neurons and the second hidden layer has  $n$  neurons. The corresponding weight vectors are  $w, w',$  and  $w''$ , and the input vectors are  $x, x',$  and  $x''$ . Suppose for  $P$  learning sample vectors, the goal of learning is to minimize the following mean square error:

$$E(W) = 0.5 \cdot \prod_{p=0}^{P-1} \prod_{i=0}^{N_M-1} [y(l,p) - d(l,p)]^2. \quad (3)$$



Among them,  $d(l, p)$  and  $y(l, p)$ , respectively, represent the target output of the  $p$ -th training mode corresponding to the  $l$ -th sample and the actual output of the network, and  $N_M$  is the number of neurons in the output layer. For simplicity, the offset threshold  $\theta$  is also incorporated into the weight vector. The error is propagated back, and the weights of each layer are corrected in turn. The output layer is

$$w''_{kl}(t+1) = w''_{kl}(t) - (1-\eta) \cdot \prod_{p=0}^{P-1} [x \cdot \delta(kl, p)], \quad (4)$$

$$\delta(kl, p) = y(l, p) \cdot [1 - y(l, p)] \cdot [d(l, p) - y(l, p)].$$

Among them,  $\eta > 0$  is called the learning rate and  $t$  is the number of iterations. The middle hidden layer is

$$w'_{jk}(t+1) = w'_{jk}(t) + \eta \cdot \prod_{p=0}^{P-1} x' \cdot \delta(jk, p). \quad (5)$$

Among them,

$$\delta(jk, p) = \prod_{l=1}^{N_M} \delta(kl, p) \cdot x(k, p) \cdot [1 - x(k, p)]. \quad (6)$$

The BP algorithm solves the learning problem of multilayer perceptrons and makes the research of neural networks enter a new stage. Of course, the BP algorithm has some shortcomings while achieving success, the most important of which is that the algorithm tends to fall into local extremes. From a mathematical point of view, BP is a gradient descent algorithm, which is very sensitive to the selection of initial values and step lengths. Since the training error surface is usually multipeak, the BP algorithm easily falls into the local optimum, and the generalization ability of the network is not strong when the network structure is determined.

Training error and prediction error (test error) are two important indicators to measure the performance of neural networks. In particular, the prediction error, which characterizes the generalization ability of the trained neural network, is a key indicator that determines whether a network can be applied. Generally speaking, the generalization ability of a neural network is determined by the topological structure and network weights. The topology design of the neural network has a lot to do with the specific form of the problem to be solved.

**4.3. Weight Optimization Model Based on IQEA-BP Neural Network.** Aiming at the problem that the BP algorithm easily falls into the local minimum, this paper proposes the IQEA-BP neural network training method. The neural network training process can be regarded as an optimization problem. By finding a set of optimal real number weight combinations, the error between the output result and the expected result under this weight is minimized. Quantum evolutionary algorithm is a better way to find this optimal weight combination.

The main idea of using the improved quantum genetic algorithm to optimize the weights of the BP neural network is as follows: First, use the quantum genetic algorithm to optimize the initial weight distribution and find a better search space in the solution space. In this search space, there are  $\sigma(\sigma \geq 1)$  which set the optimal initial weight combination to overcome the shortcomings of the BP algorithm wherein it is easy to fall into the local optimum and is more sensitive to the initial value setting; then, use the BP algorithm gradient descent principle to further "fine-tune" the weights; the optimal solution is searched out in this smaller solution space, which is the true global optimal point. Using the improved quantum genetic algorithm to optimize the weights of the neural network can better prevent the search from falling into the local minimum.

For the optimal solution obtained, the weights and thresholds of the  $\sigma$  group neural network with the smallest error are used as the initial weights of the neural network for training. If the accuracy requirements cannot be met, the neural network training can be performed again. The convergence speed of the algorithm is accelerated by introducing a mechanism for dynamically adjusting the rotation angle of the quantum gate. By adopting quantum crossover and quantum mutation operations, the interference and entanglement characteristics of quantum states are fully utilized, so the algorithm has better optimization performance and can effectively prevent premature maturity.

Because it has gone through the global search of the improved quantum evolution algorithm, it is unlikely that it will fall into a local minimum again. The training is repeated until a set of neural network weights and thresholds that meet the accuracy requirements are obtained. The schematic diagram of the entire algorithm is shown in Figure 4.

## 5. Experimental Results and Analysis

**5.1. Forward and Standardization of Reverse Indicators.** In order to protect school information, all school names involved in this article use virtual names. The evaluation index data of some schools is shown in Figure 5.

Before using analytical methods to comprehensively evaluate multiple indicators, it is necessary to judge whether the original indicator variables have reverse indicators. If there is a reverse index, it needs to be normalized so that all the evaluation indexes tend to the same direction. Analyzing the data characteristics of the index variables found that they are all positive index data, so there is no need to carry out the normalization process.

**5.2. Calculate the Covariance Matrix.** The covariance matrix of the standardized data eliminates the influence of the variance of a single indicator variable and retains the correlation between the indicator variables. The index with strong correlation and a large number of related indexes in the covariance matrix is used as the principal component. This principal component has a large overlap with other indexes and contains more information about the original data. The covariance matrix of the standardized experimental data is shown in Table 1.

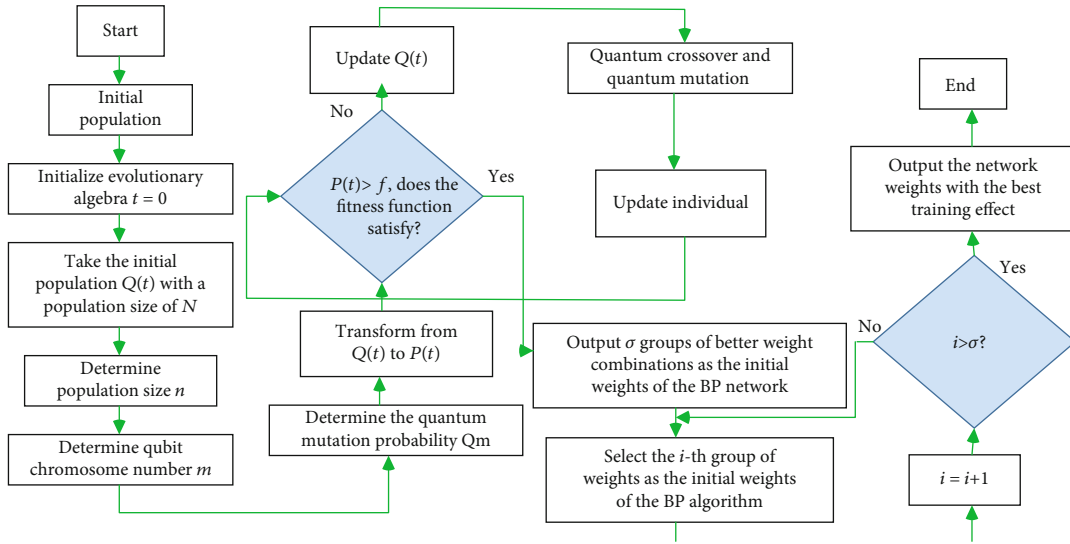


FIGURE 4: Flow chart of IQEA-BP algorithm.

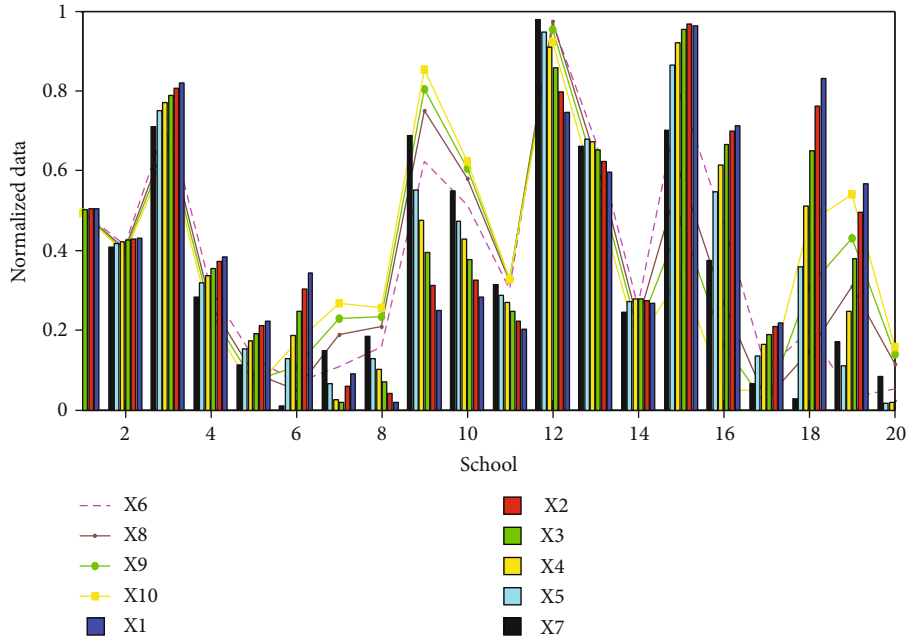


FIGURE 5: Data of some schools in a city.

In Table 1, the matrix of X1-X9 represents the covariance matrix of the old evaluation index, and the matrix of X1-X10 represents the covariance matrix of the new evaluation index. From the data in Table 1, it can be concluded that most of the variables have a high correlation. For example, the total number of students has a high correlation with the other 10 evaluation indicators, which proves that there is overlap of information between the new and old evaluation indicators.

**5.3. Results of Comprehensive Evaluation of Education Quality.** Through the positive and negative analysis of the evaluation indicators, it can be known that X1-X10 are all positive indicators, that is, the larger the indicator data, the better. We analyze these evaluation indicators to get a com-

prehensive ranking of school education quality. The smaller the ranking value, the better the school education quality. Figure 6 shows the distribution of the new evaluation indicators X1-X10 after the school's education quality evaluation ranking. It can be seen from the figure that there are only individual indicator data that do not match the school's ranking results. The change trend of most of the new evaluation indicators X1-X10 is consistent with that of the school's ranking. The school ranking results are consistent, that is, the higher the data of the new evaluation index X1-X10, the better the school education quality and the higher the school ranking; the lower the data of the new evaluation index X1-X10, the worse the school education quality and the higher the school ranking lean back (see Figure 6).

TABLE 1: Covariance matrix.

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
X1	0.98	0.87	0.82	0.85	0.83	0.87	0.82	0.90	0.76	0.82
X2	0.71	1	0.87	0.82	0.82	0.81	0.97	0.86	0.87	0.89
X3	0.71	0.71	1	0.87	0.82	0.86	0.82	0.87	0.82	0.83
X4	0.83	0.71	0.71	0.99	0.87	0.82	0.89	0.87	0.82	0.85
X5	0.87	0.82	0.89	0.81	0.99	0.87	0.82	0.8	0.71	0.93
X6	0.87	0.82	0.87	0.83	0.75	1	0.87	0.82	0.8	0.91
X7	0.87	0.82	0.84	0.83	0.73	0.76	0.98	0.87	0.82	0.88
X8	0.87	0.82	0.84	0.84	0.87	0.82	0.89	1	0.71	0.77
X9	0.87	0.82	0.86	0.71	0.87	0.82	0.83	0.88	1	0.91
X10	0.87	0.82	0.87	0.82	0.87	0.82	0.84	0.87	0.75	0.99

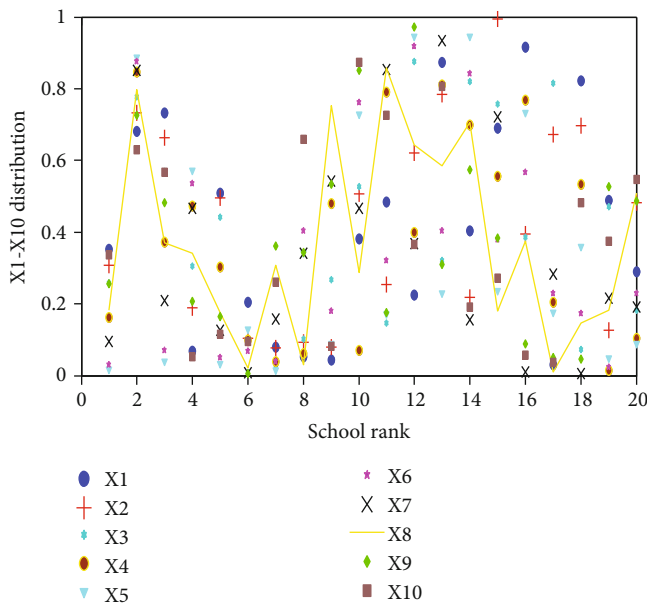


FIGURE 6: Distribution of X1-X10 after school ranking.

In order to quantitatively compare the evaluation effect of analysis methods, this paper uses the accuracy rate to analyze the evaluation effect of the IQEA-BP neural network. The results of the evaluation index experiment are shown in Figure 7.

It can be seen from Figure 7 that the accuracy of the IQEA-BP neural network is higher than that of the BP neural network. It proves that the evaluation result of the IQEA-BP neural network is closer to the empirical value of school education quality grade gradient. Therefore, this article uses the IQEA-BP neural network as the evaluation method of the school education quality evaluation system. After the evaluation index data X1-X10 are trained by the neural network for 3000 times, the error rate of the training set and the test set is obtained, as shown in Figure 8. It can be seen from Figure 8 that the error rate of the IQEA-BP neural network training set and the test set is very low, and the effect of the school education quality evaluation model of the IQEA-BP neural network can be obtained generally subjectively (see Figure 8).

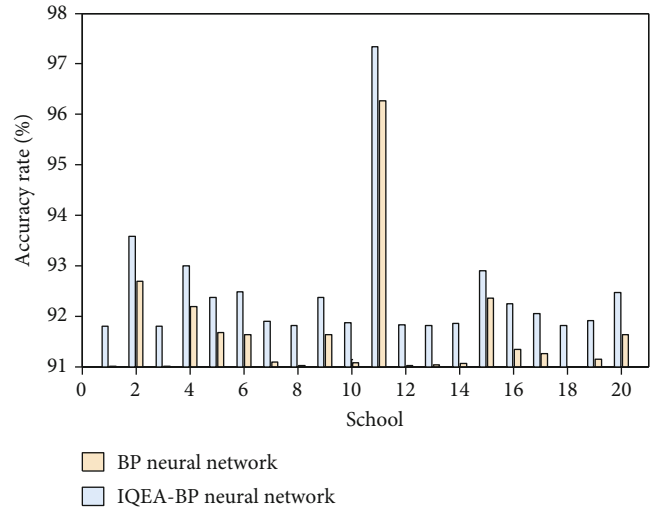


FIGURE 7: Comparison of experimental results of accuracy evaluation.

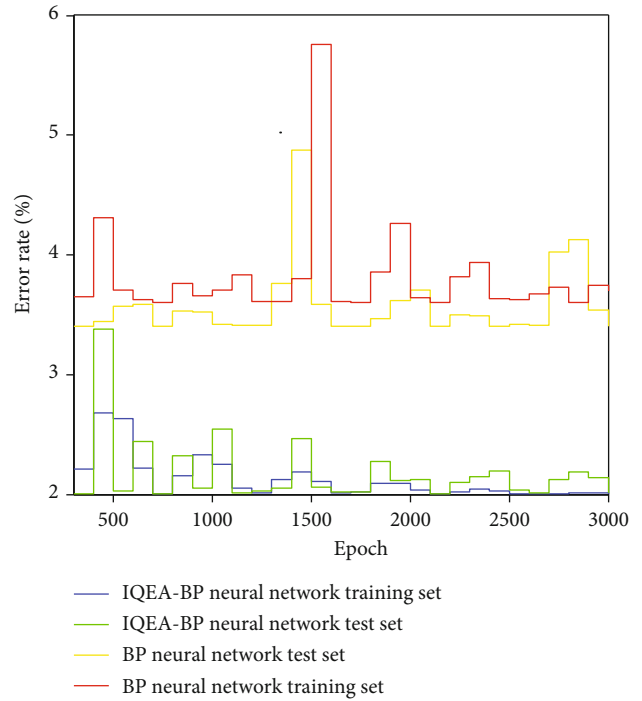


FIGURE 8: Epoch = 3000 experimental error rates.

In order to clearly and objectively analyze the effect of the neural network evaluation model, the accuracy and recall rate are used to test the evaluation effect of the model. Accuracy refers to the proportion of the number of samples that are truly positive among all predicted positive samples. The recall rate is based on the true value, the proportion of all positive samples that are predicted to be the number of positive samples.

It can be seen from Figure 9 that the time-consuming assessment of the educational quality of 20 schools is concentrated in 2 s to 5 s, and the time required for the IQEA-BP

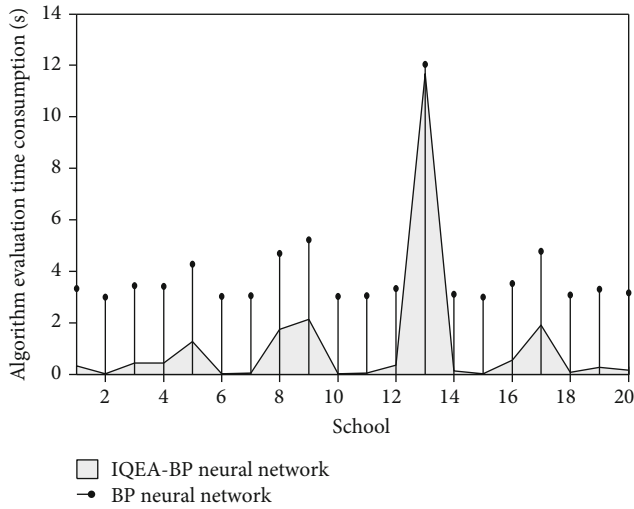


FIGURE 9: The time-consuming evaluation of education quality in 20 schools.

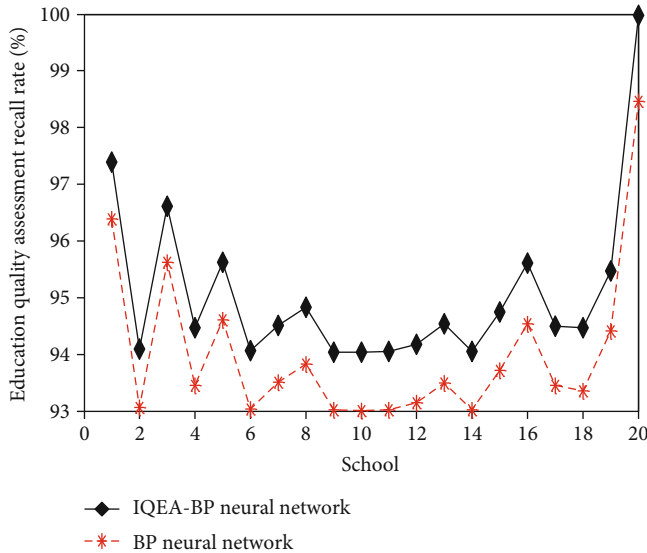


FIGURE 10: Education quality assessment recall rate of 20 schools.

neural network assessment is lower. It can be seen from Figure 10 that the recall rate of the education quality evaluation of 20 schools is above 93%, which shows that the school education quality evaluation model of the IQEA-BP neural network in this paper is feasible. Therefore, the school education quality evaluation model of the IQEA-BP neural network has achieved a good evaluation effect, and the evaluation effect is more ideal than that of the BP neural network (see Figures 9 and 10).

## 6. Conclusion

In the context of the education cloud platform, this topic analyzes the uneven distribution of education application service resources, low resource utilization, and repeated development of service tools faced in the process of education infor-

matization and elaborates the construction of a reasonable application service architecture. According to the functional requirements and characteristics of the education cloud platform application service, combined with the use of software and services, an architecture model of the application service based on the education cloud platform is proposed. This paper proposes a BP neural network weight optimization model based on an improved quantum evolution method and studies a method that combines the improved quantum evolution algorithm (IQEA) with the BP algorithm to complete the back propagation neural network training. The algorithm improves the traditional quantum evolution algorithm and then uses the improved quantum evolution algorithm to optimize the network weights as a whole, to overcome the shortcomings of the BP algorithm that easily falls into the local optimum. The BP algorithm is further optimized to improve the training and prediction accuracy of the network. The analysis method is used to comprehensively evaluate the school education quality of the old and new two groups with different evaluation indexes, and the evaluation results are compared and analyzed to determine the selection of evaluation indexes. A school education quality evaluation model based on the BP neural network's compulsory education equilibrium index is established. The evaluation model is a classification problem, and the experimental data is processed with uniform distribution and normalization. The design of the BP neural network evaluation model includes the determination of the network structure, the selection of the appropriate activation function, the determination of the loss function, and the selection of the optimizer. At the same time, we select the appropriate activation function and network structure to solve the BP neural network under-fitting problem, construct a suitable network structure, and use the Dropout idea to solve the BP neural network over-fitting problem.

## Data Availability

All information is within the paper.

## Conflicts of Interest

No competing interests exist concerning this study.

## Acknowledgments

This work is partially supported by the Shaanxi Higher Education Teaching Reform Research Project (19ZY016) and Shaanxi Education Scientific Planning Project in 2016 (SGH16H229).

## References

- [1] Y. Zhu, H. Lu, P. Qiu, K. Shi, J. Chambua, and Z. Niu, "Heterogeneous teaching evaluation network based offline course recommendation with graph learning and tensor factorization," *Neurocomputing*, vol. 415, pp. 84–95, 2020.
- [2] H. Xue, Y. Lin, Y. Yuan, and J. Cai, "Early warning classification of cluster supply chain emergency based on cloud model

- and datastream clustering algorithm,” *Journal of Intelligent & Fuzzy Systems*, vol. 35, no. 1, pp. 393–403, 2018.
- [3] L. Y. Leong, T. S. Hew, K. B. Ooi, and A. Y. L. Chong, “Predicting the antecedents of trust in social commerce - a hybrid structural equation modeling with neural network approach,” *Journal of Business Research*, vol. 110, pp. 24–40, 2020.
- [4] B. Rajalingam and R. Priya, “A novel approach for multimodal medical image fusion using hybrid fusion algorithms for disease analysis,” *International Journal of Pure and Applied Mathematics*, vol. 117, no. 15, pp. 599–619, 2017.
- [5] X. Jiang and Y. D. Zhang, “Chinese sign language fingerspelling via six-layer convolutional neural network with leaky rectified linear units for therapy and rehabilitation,” *Journal of Medical Imaging and Health Informatics*, vol. 9, no. 9, pp. 2031–2090, 2019.
- [6] X. Zhang, C. Fu, Y. Zhao, and X. Xu, “Hybrid feature CNN model for point cloud classification and segmentation,” *IET Image Processing*, vol. 14, no. 16, pp. 4086–4091, 2020.
- [7] M. Chiregi and N. J. Navimipour, “Cloud computing and trust evaluation: a systematic literature review of the state-of-the-art mechanisms,” *Journal of Electrical Systems and Information Technology*, vol. 5, no. 3, pp. 608–622, 2018.
- [8] F. Moslehi and A. Haeri, “A novel hybrid wrapper-filter approach based on genetic algorithm, particle swarm optimization for feature subset selection [J],” *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, no. 3, pp. 1105–1127, 2020.
- [9] Z. Yu, A. M. Abdulghani, A. Zahid, H. Heidari, M. A. Imran, and Q. H. Abbasi, “An overview of neuromorphic computing for artificial intelligence enabled hardware-based Hopfield neural network,” *IEEE Access*, vol. 8, pp. 67085–67099, 2020.
- [10] B. Wang, S. J. Hu, L. Sun, and T. Freiheit, “Intelligent welding system technologies: State-of-the-art review and perspectives,” *Journal of Manufacturing Systems*, vol. 56, pp. 373–391, 2020.
- [11] W. Liu, H. L. Ma, and A. Walsh, “Advance in photonic crystal solar cells,” *Renewable and Sustainable Energy Reviews*, vol. 116, p. 109436, 2019.
- [12] X. Zhang, C. Zang, H. L. Ma, and Z. J. Wang, “Study on removing calcium carbonate plug from near wellbore by high-power ultrasonic treatment,” *Ultrasonics Sonochemistry*, vol. 62, p. 104515, 2020.
- [13] H. L. Ma, X. Zhang, F. F. Ju, and S. B. Tsai, “A study on curing kinetics of nano-phase modified epoxy resin,” *Scientific Reports*, vol. 8, no. 1, p. 3045, 2018.
- [14] M. Ling, M. J. Esfahani, H. Akbari, and A. Foroughi, “Effects of residence time and heating rate on gasification of petroleum residue,” *Petroleum Science and Technology*, vol. 34, no. 22, pp. 1837–1840, 2016.
- [15] H. L. Ma and S. B. Tsai, “Design of research on p of a new iridium coordination compound for the detection of Hg<sup>2+</sup>,” *International Journal of Environmental Research and Public Health*, vol. 14, no. 10, p. 1232, 2017.
- [16] L. Y. Mo, W. H. Z. Sun, S. Jiang et al., “Removal of colloidal precipitation plugging with high-power ultrasound,” *Ultrasonics Sonochemistry*, vol. 69, p. 105259, 2020.
- [17] D. Gao, Y. Liu, Z. Guo et al., “A study on optimization of CBM water drainage by well-test deconvolution in the early development stage,” *Water*, vol. 10, no. 7, p. 929, 2018.
- [18] S. B. Tsai and H. A. Ma, “Research on preparation and application of the monolithic catalyst with interconnecting pore structure,” *Scientific Reports*, vol. 8, no. 1, 2018.
- [19] J. Xie and H. Ma, “Application of improved APO algorithm in vulnerability assessment and reconstruction of microgrid,” *Iop Conference*, vol. 108, no. 5, article 052109, 2018.
- [20] A. Fujin, Y. Xiuzhao, H. Ruochi, M. Hailing, and L. Wei, “Research into the super-absorbent polymers on agricultural water,” *ScienceDirect. Agric Water Management.*, vol. 245, p. 106513, 2021.
- [21] Y. Zhang, Z. Zhou, H. Bai, W. Liu, and L. Wang, “Seizure classification from EEG signals using an online selective transfer TSK fuzzy classifier with joint distribution adaption and manifold regularization,” *Frontiers in Neuroscience*, vol. 14, pp. 1–9, 2020.
- [22] Y. Zhang, F. L. Chung, and S. Wang, “Fast exemplar-based clustering by gravity enrichment between data objects,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 50, no. 8, pp. 2996–3009, 2020.
- [23] Y. Zhang, F. L. Chung, and S. Wang, “Takagi-Sugeno-Kang fuzzy systems with dynamic rule weights,” *Journal of Intelligent & Fuzzy Systems*, vol. 37, no. 6, pp. 8535–8550, 2019.
- [24] T. Grubljesic, P. S. Coelho, and J. Jaklic, “The shift to socio-organizational drivers of business intelligence and analytics acceptance,” *Journal of Organizational and End User Computing*, vol. 31, no. 2, pp. 37–64, 2019.
- [25] L. X. Z. Zhang, M. Mouritsen, and J. R. Miller, “Role of perceived value in acceptance of “bring your own device” policy,” *Journal of Organizational and End User Computing*, vol. 31, no. 2, pp. 65–82, 2019.
- [26] A. Shahri, M. Hosseini, K. Phalp, J. Taylor, and R. Ali, “How to engineer gamification: the consensus, the best practice and the grey areas,” *Journal of Organizational and End User Computing*, vol. 31, no. 1, pp. 39–60, 2019.

## Research Article

# An Adaptive BP Neural Network Model for Teaching Quality Evaluation in Colleges and Universities

Yong Jin <sup>1</sup>, Yiwen Yang,<sup>2</sup> Baican Yang,<sup>2</sup> and Yunfu Zhang<sup>3</sup>

<sup>1</sup>Department of Academic Affairs, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China

<sup>2</sup>Department of Pharmacy, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China

<sup>3</sup>Department of Organization, Yunnan University of Traditional Chinese Medicine, Kunming 650000, China

Correspondence should be addressed to Yong Jin; [jinyong@shutcm.edu.cn](mailto:jinyong@shutcm.edu.cn)

Received 20 June 2021; Revised 17 July 2021; Accepted 26 July 2021; Published 15 August 2021

Academic Editor: Shan Zhong

Copyright © 2021 Yong Jin et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

There is currently no fair, rational, or scientific approach for evaluating college teachers' teaching abilities. Mathematical methods are frequently used to measure the teaching capacity of college instructors in order to make it more scientific. Traditional statistical analysis evaluation models, fuzzy evaluation methods, grey decision methods, and the analytic hierarchy process (AHP) are only a few examples. Because teacher assessment is a nonlinear problem, even though the preceding methods have produced some positive results, they are vulnerable to some subjectivity. In this paper, the neural network model is incorporated into the adaptive vector and momentum of the modified BP neural network of a gradient descent method to boost the model's convergence speed, and the model is thoroughly researched to evaluate university teaching quality, and the network structure is omitted to address the complex nonlinear problem of college and university teaching quality assessment. The model's comprehensive evaluation of teaching activities is then bolstered by the addition of new evaluation indexes to the existing ones.

## 1. Introduction

With the rapid development of higher education [1–4], a school's reputation has become the most important criterion for students when choosing a school, and a school's reputation is largely determined by the quality of its education [5–7]. A school's top priority is education quality, because it has an impact on not only the school's existence and development but also the future and destiny of its students. However, as a result of the large-scale enrollment increase of schools and universities for several years in a row, a number of related issues have emerged, including a teacher shortage, a decline in the quality of students, and a shortage of educational and teaching equipment and logistics facilities. These issues have sparked widespread concern in society, prompting a discussion about how to address them. Education quality has become a contentious issue, but it has also evolved into a comprehensive reflection of college and university work.

Improving educational quality is always a hot topic in higher education. The idea of "improving higher education

quality" is very practical. As a result, the internal assessment activity places a strong emphasis on the creation of a system for evaluating teaching quality. The teaching quality of colleges and universities is reflected in the teaching quality of each specialty or department, the teaching quality of each department is reflected in the quality of each course, and the course quality is reflected in the teaching quality of each teacher who teaches the course [8, 9]. The quality and degree of teachers have an impact on how a school is run. The quality and level of teachers determine the administrative direction of a school, the quality of education and instruction, and the impact of educational reform.

Teaching is the most important aspect of schoolwork. The effectiveness of a school's teaching is a key determinant of its success [10]. Teaching quality management is critical to the school's overall quality management. Evaluation of teachers' teaching quality aids school leaders and administrators in determining the degree to which teaching goals have been met, grasping the school's teaching work comprehensively and accurately, and improving teaching quality. The topic of teaching quality evaluation [11–13] will also be

discussed. This is a critical assignment. The cornerstone of teaching activities in colleges and universities, as well as the most significant connection in assuring the quality of talent development in these institutions, is teaching quality evaluation. As a result, assessing teaching quality is a critical component of promoting high-quality education, changing teaching, and improving teaching quality. It has a substantial impact on overall educational quality. Because teaching is both a spiritual labor and an art, there is no established formula. Nonquantitative criteria are widely utilized in the evaluation of teaching quality, because defining what constitutes teaching is very ambiguous and difficult to measure. The difficulty and complication of assessing quality teaching and learning are both part of the teaching process. It is far more difficult to evaluate the quality of a teacher's instruction than it is to evaluate the quality of a product. The teaching process is made up of a variety of components, and it is a two-way activity that involves both teachers and students. There are many factors that influence teaching quality, and there is still more to investigate. The development of a scientific and reasonable teaching quality evaluation system [14] that can accurately and equitably assess teaching quality is a significant challenge.

Because it is a complex and abstract nonlinear problem, it is difficult to express the evaluation of teaching quality at colleges and universities using a mathematical model or analytical formula. The neural network model [15–17] can achieve mutual mapping between any dimensions and has strong nonlinear processing capabilities [18, 19]. As a result, creating a neural network model to handle the challenge of evaluating the quality of college instruction is an effective metric. Computer vision [20–22] in the evaluation of teaching quality in colleges and universities not only aids in improving teaching quality, promoting continuous improvement of teaching goals, and promoting scientific education decision-making, but it also aids in the development of intelligent and standardized teaching management in colleges and universities.

The following are the main innovations points of this paper:

- (1) This study presents a novel adaptive BP neural network model that can adequately evaluate the teaching quality and teaching effect of colleges and universities in order to properly evaluate the teaching quality and teaching effect of colleges and universities
- (2) As a novel adaptive BP neural network, a model is proposed in this study. The model contains an adjustable learning rate and momentum term to improve the gradient descent method of the BP neural network's convergence speed and optimize the network topology to ensure the model's stability

## 2. Background

Teaching quality evaluation in college is a complicated topic that encompasses variables such as teaching conditions, class difficulty, instructor teaching and learning effect of many

aspects, such as their interaction, and the complex relationship between teachers and students at the same time; factors affecting the quality of teaching are also more; at present, none of them has been recognized; the ideal teaching quality evaluation system, in the context of existing research status, focuses on three aspects: the first is the study of the evaluation subject, the second is an investigation into the content of the teaching quality assessment system, and the third is an investigation on how to evaluate the teaching quality grading technique once each index in the system has been determined.

Teachers' self-evaluation, peer evaluation [23], administrative leadership evaluation, expert assessment, and student evaluations of teachers are all examples of ways or techniques to assess teaching quality. Because each evaluation method and its outcomes serve a different purpose in the evaluation, their evaluation functions should be distinct as well. Each evaluation method and its results are only a part of the overall assessment of teaching quality, but they cannot be compared. Because of the large number of college and university teachers, as well as the frequent evaluation of the number of teachers, the organization's leadership and peer expert survey evaluation method is not only time-consuming. Because of the interpersonal contact and lack of familiarity with the teaching process, as well as the influence of such aspects as actually difficult to operate, most colleges and universities use students as the major body of teachers' teaching quality evaluation technique. China's colleges and universities have used student evaluation of instructional activities since the 1980s, which has helped to improve teaching quality. Because they are the direct recipients of instruction, students have the right and ability to evaluate teachers' teaching. Because of the variety of types, complex majors, and uneven levels of students, different requirements for instructors' teaching quality evaluation exist in colleges and universities across the country.

On the content of the teaching level evaluation system design, it is difficult to take a course and a teacher in a learning phase, to quantify the effect of general courses as the main index, or teaching effect as the main index, and put the evaluation content on the teaching process, because learning and development is a continuous process and learning and growth environment is diverse. From the perspective of process management, the school teaching process is manifested as the interaction of multiple factors and the combination of multiple links; it is also difficult to compare different fields' teaching, different types of courses, teaching relationships, and teaching objects. As a result, the most basic features that can immediately represent the teaching level and are prevalent are mostly taken into account when evaluating teaching quality.

## 3. Methodology

*3.1. Teaching Quality Evaluation System.* The goal of teaching quality assessment is to promote teaching reform, improve teaching quality, reduce student burden, increase students' intelligence, and improve their ability to analyze and solve problems. We should be objective, fair, and rational

when evaluating the quality of instruction, rather than guessing or mixing personal feelings, and we must achieve the unity of ideology, science, and feasibility. Teaching quality is frequently assessed in colleges and universities through four channels: student evaluation, expert evaluation, peer evaluation, and instructor self-assessment, with the final evaluation results synthesized. However, various concerns remain in the process of establishing, employing, and evaluating the teaching quality assessment system's evaluation outcomes, such as evaluation theory research, evaluation method use, evaluation method updates, and evaluation data analysis. These issues have an immediate impact on the educational system. The importance of quality assurance and the potential for knowledge mining.

The various indicators in the evaluation system generally involve teaching attitudes, teaching content proficiency, and basic teaching skills. However, the comprehensive quality of teachers is not only reflected in the above aspects but should also include teachers' knowledge, teaching and research ability, teaching design ability, and teacher's innovation ability, but these evaluation indicators that can fully reflect the comprehensive quality and personality of teachers are rarely involved in the evaluation system. As a result, the design of an evaluation index system should be given careful thought.

The indications in the evaluation system of the differences in the influence degree of the evaluation findings should be given distinct weights; however, many schools and universities still use the same weighting approach to make things easier for themselves or subjectively determined a weight distribution table to establish the evaluation system; using this evaluation approach to assess teaching quality not only undermines the credibility of evaluation outcomes but also creates barriers to further data mining. As a result, a suitable weight distribution is a critical step in perfecting the evaluation system.

The data in the teaching evaluation database is continuously growing as network teaching assessment becomes more prevalent on digital campuses. Facing the "mountain of data sets," traditional data analysis means can only obtain the surface information of these data, and it is difficult to dig out the deep valuable information. An essential goal of teaching evaluation is to learn how to better use evaluation data to extract knowledge. The lack of modern scientific and technology methodologies and an imperfect evaluation feedback and control mechanism are the major issues that the teaching quality evaluation system is currently experiencing. The correctness and trustworthiness of evaluation outcomes will be directly affected by the scientific establishment of assessment indexes and the rationality of index weight distribution. As a result, developing and establishing a method for evaluating teaching quality are a vital step in enhancing teacher quality and a pressing requirement in colleges and universities to improve classroom management.

Teachers' teaching quality is evaluated in terms of content, techniques, attitude, and teaching effect, such as content, in order to demonstrate consistency, comprehensiveness, and effectiveness, and the principle of independence, incentive, and fault tolerance; this article set

evaluation index is from 1 to 7:  $X_1$ ,  $X_2$ ,  $X_3$ ,  $X_4$ ,  $X_5$ ,  $X_6$ , and  $X_7$ , where  $X_1$  represents the course progress, the depth and breadth of teaching, and the reasonable degree of learning burden;  $X_2$  represents the degree of integration with practice and whether it can reflect modern scientific and technological achievements;  $X_3$  represents the degree of clarity, hierarchy, and emphasis of the lecture;  $X_4$  stands for vivid explanation, inspiring and inducing, and attractive, exemplified by examples, combining theory with practice;  $X_5$  stands for instructing correct learning methods to cultivate students' analytical ability;  $X_6$  represents the degree of careful preparation for lessons, skilled explanation, answering questions, and correcting homework; and  $X_7$  stands for continuous improvement in teaching and educating people.

*3.2. Adaptive BP Neural Network.* Existing teaching quality evaluation methods and models in colleges and universities have flaws in processing small-scale low-dimensional data sets, such as difficulty determining the index weights of the analytic hierarchy process, subjectivity and randomness of the fuzzy synthesis method, and the support vector machine's grid search. When using a traditional BP neural network, the optimization approach is slow, the self-convergence is slow, and it is easy to fall into the local minimum. An adaptive BP neural network model is proposed in this study. Integrating an adaptive learning rate and a momentum component to speed up convergence, as well as tuning the network structure to ensure model stability, improves the gradient descent methodology of the BP neural network. New evaluation indicators are added to the model's input feature vector to ensure a complete assessment of teaching activities, and the evaluation sample data set is normalized to improve the model's computation performance. The evaluation sample data set's training and test portions are separated and entered into the model for training and verification. To ensure that the model presented in this chapter is useful in assessing teaching quality at colleges and universities, performance indicators such as MSE, prediction accuracy, and training time are used to compare it to other approaches and models.

Given the fact that assessing teaching quality is a multi-objective, multilevel, and sophisticated nonlinear problem and current college teaching quality assessment methods and models make determining weights difficult, subjectivity and randomness are too strong, and the conventional BP neural network has a slow convergence time, making it easy to fall into a local minimum. The model of an adaptive BP neural network is proposed. The basic idea behind this model is to use an adaptive learning rate and a momentum term to improve the convergence speed of a BP neural network's gradient descent method and optimize the network topology to ensure the model's stability. In addition, to create the teaching quality evaluation index system of cost text, new evaluation indexes were added to the traditional evaluation indexes, ensuring the model's comprehensive evaluation of teaching activities, and the evaluation index sample data set was normalized as the model's input feature vector, improving the model's calculation efficiency.



**3.2.1. Network Structure.** Any nonlinear mapping of  $n$ -dimensional to  $m$ -dimensional functions can be satisfied by a three-layer BP neural network. As a result, this research develops a three-layer BP neural network model for evaluating teaching quality, with one input layer, one hidden layer, and one output layer. The size of the input sample and the size of the output result, respectively, determine the number of neurons in the input and output layers. Selecting neurons in the buried layer, on the other hand, is more difficult and not well guided by theory. This research uses an empirical formula and multiple experiments to identify the number of neurons in the input and output layers. Using equation (1) and trial and error, we establish the appropriate number of neurons.

$$n_i = \sqrt{n + m} + a, \quad (1)$$

where the number of neurons in the input layer is  $n$ , the number of neurons in the output layer is  $m$ , and  $a$  is a constant between [1, 10].

**3.2.2. Adaptive Learning Rate and Momentum.** In a standard BP neural network, the learning rate, also known as the learning step size, is fixed. When the learning rate is too high, the network topology becomes unstable and oscillates. However, if the learning rate is too low, the network's convergence speed will be slow, making it impossible to ensure that the learning efficiency of the overall network structure is optimal in real-world applications. To increase convergence speed, the adaptive learning rate automatically adjusts the learning rate in response to changes in network error and corrects the weights and thresholds between the connection layers on a regular basis. Assuming the initial learning rate is  $\mu(0)$ , the  $n$ -th iteration of the model results in a network error of  $E(n)$ . When the learning rate changes, the equation is shown as follows:

$$\mu(n) = \begin{cases} \beta\mu(n-1), & E(n) < E(n-1), \\ \gamma\mu(n-1), & E(n) > E(n-1), \\ \mu(n-1), & \text{else,} \end{cases} \quad (2)$$

where  $\beta = 1.05$  and  $\gamma = 0.7$ .

Adaptively modifying the learning rate throughout the error back propagation process can significantly improve convergence speed. However, only the adjustment of the gradient descent direction at the current time  $t$  is considered, and the gradient direction before the time  $t$  is not considered, which will cause turbulence in the training process and cause the model to become unstable and easily fall into a local minimum. The momentum term is inserted to rectify this inconsistency, and it alters the weight value to give a damping effect in the error back propagation process. The equation is shown in (3). Regarding equation (3) as a time series with  $t$  as a variable ( $0 < t < N$ ), equation (3) can be regarded as the first-order difference equation of  $\Delta w(n)$ , and its calculation equation is as follows:

$$\Delta w(n) = \alpha \Delta w(n-1) - \mu \frac{\partial E(n)}{\partial w(n)}, \quad (3)$$

$$\Delta w(n) = -\mu \sum_{t=0}^n \alpha^{n-t} \frac{\partial E(n)}{\partial w(n)}, \quad (4)$$

where  $\alpha$  is the momentum term ( $0 < \alpha < 1$ ),  $w$  is the weight,  $\mu$  is the learning rate, and  $E(n)$  is the error. Then, the weight adjustment equation of BP neural network is as follows:

$$w(n+1) = w(n) - \mu(n) \sum_{t=0}^N \alpha^{n-t} \frac{\partial E(n)}{\partial w(n)}. \quad (5)$$

To ensure a more thorough evaluation of the teacher's teaching process, 23 second-level indicators were produced by merging two first-level indicators of preteaching preparation and the situation in the teaching process with standard evaluation indicators in this study. Normalize the evaluation sample data to reduce the difficulty of altering the weight and threshold due to the large change in the input value, which will increase the calculation efficiency of the BP neural network.

**3.3. Evaluation of University Teaching Quality Based on the Adaptive BP Neural Network.** Figure 1 shows the processing flow chart for the adaptive BP neural network model for teaching quality evaluation. To increase the model's convergence speed, the model adjusts the learning rate based on a comparison of the current error and the prior error and adds a momentum term to dampen the model to prevent oscillations and assure stability. At the same time, new input variables are introduced. To ensure a fuller model evaluation, the input features are included. From the normalized evaluation index sample number set, a training data set and a test data set are constructed, with the training data set being used to train the model. Adjust the number of hidden layer neurons in the network structure of the model during the training process, using the adaptive learning rate and momentum term to improve the gradient descent method of the BP neural network as the training function. The rate and momentum terms are fine-tuned, and the model was continuously trained iteratively to modify its own weights and thresholds until the evaluation result meets the set target accuracy or reaches the maximum number of iterations, then exits the training and obtains this time the best or final model. Finally, use the test data set to demonstrate that the proposed model has good evaluation prediction accuracy and benefits in evaluating teaching quality.

## 4. Experiments and Results

**4.1. Experimental Setup.** In the experiment, the network structure of the model is given as three layers, the learning rate is set to 0.001, the maximum number of training times is 10000, the weights and thresholds of the model are randomly selected, and the activation function is a single sigmoid function.

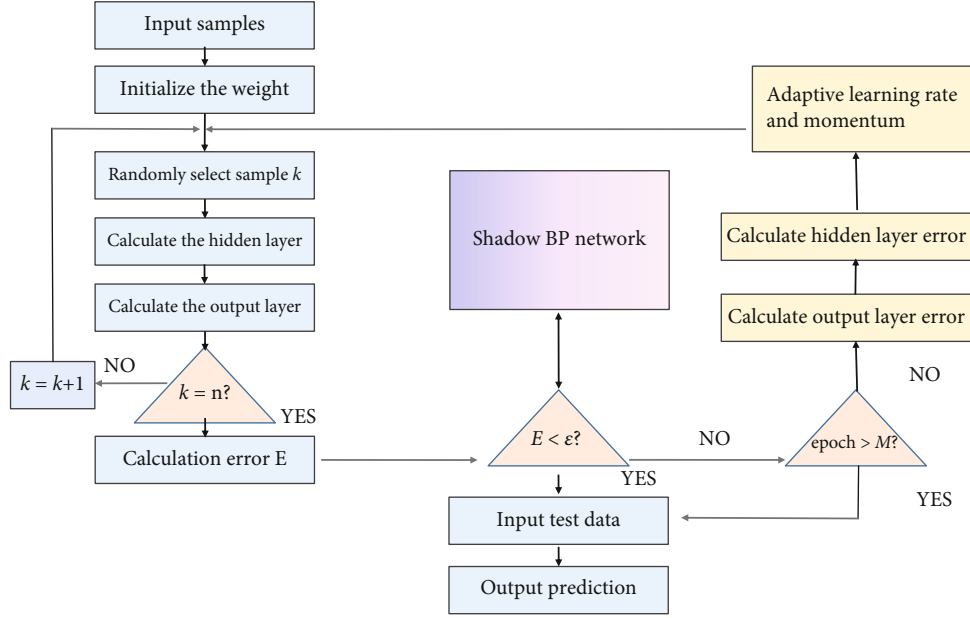


FIGURE 1: Evaluation model of teaching quality in colleges and universities.

**4.2. Dataset.** Obtain the evaluation data set of a course from 2003 to 2020 from the educational administration system of a university. There are a total of 7171 sample data points. One of them is the data used to evaluate the teacher's teaching process using students as the major evaluation body; the data used to evaluate the teacher's teaching process using students as the primary evaluation body is used as the evaluation model's input value; because it is crucial to the model's verification, the target expected output value is incorporated based on the evaluation of the teaching supervision group teachers' numerous lecture records. The model's target predicted output value is based on the score. During the study, the sample data with a high evaluation, a bad assessment, or data that was inconsistent with the facts were removed from the data set, leaving 6,178 sample data. Sample training data is critical in neural networks. Normalize the input samples to ensure that the data is within the specified range  $[0, 1]$ , which is conducive to data processing and improves network efficiency. The normalization processing equation of the input sample is

$$X = \frac{T - T_{\min}}{T_{\max} - T_{\min}}. \quad (6)$$

**4.3. Experimental Results.** The improved gradient descent methodology is more effective than the standard gradient descent strategy in finding the adaptive learning rate and momentum term as the training function of a BP neural network. Assuming that the number of neurons in the hidden layer is 10, the growth ratio and decline ratio of the adaptive learning rate are set as 1.2 and 0.7, respectively, and the momentum term is set as 0.9. Input a training data set into the model for training in order to obtain stable network model, and then an input validation test data set, respectively,

TABLE 1: Comparison of the results of the experiment.

Method	MSE	ACC
Genetic algorithm	48.32	0.7958
BP	46.99	0.8478
Adaptive BP (ours)	32.14	0.9120

in two different algorithms as the training function, to contrast the performance of the model number of iterations and the training time, the MSE and model prediction accuracy comparison of the differences between predicted results and actual results. In this experiment, the average value obtained by running 10 times is the final result, and the result is shown in Table 1. The results show that when the adaptive learning rate and the gradient descent method with improved momentum term are used as the model's training functions, the learning rate of the model adapts adaptively and dynamically in response to changes in errors, ensuring that the model's convergence speed is improved during the training process, and the error between the model and the expected value of the target can be greatly reduced. It is also more effective than traditional methods in predicting accuracy and iteration times.

Figures 2 and 3 show that when the number of iterations grows, the MSE and prediction accuracy of the teaching evaluation model jump and fluctuate in the form of fluctuations, with the MSE increasing and then decreasing, but the ACC increasing all the while. Figure 4 shows that although the number of iterations and training duration of the model is not optimal when the number of neurons in the hidden layer is 30, the difference is not significant when compared to the number of other neurons and can

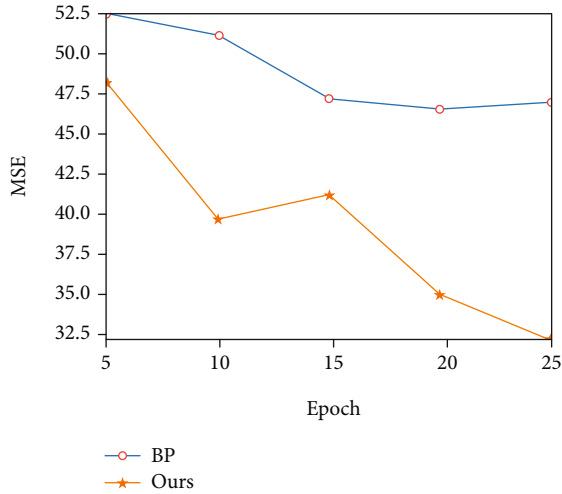


FIGURE 2: Change curve of convergence accuracy.

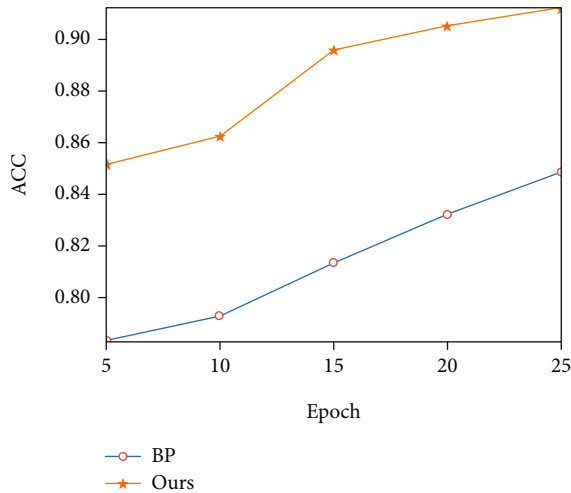


FIGURE 3: Change curve of ACC.

be tolerated; and at this time, the model of the MSE error and prediction accuracy rate are the best, which is more conducive to accurately evaluating teaching activities. As a result, based on the overall performance measures, the number of hidden layer neurons chosen is 30. Furthermore, this demonstrates that the method presented in this research can accurately assess the teaching quality of colleges and institutions.

**4.4. Ablation Experiment.** In order to further verify the effectiveness of the algorithm in this paper, we conducted ablation experiments to verify the influence of the BP network and multilayer perceptron (MLP) on the experimental results. The experimental results are shown in Table 2.

It can be seen from the experimental results in Table 2 that the performance of using MLP is far worse than that of using the BP network selected in this paper. Therefore, the MLP without optimized training is inferior to BP. Therefore, the proposed method uses the BP network.

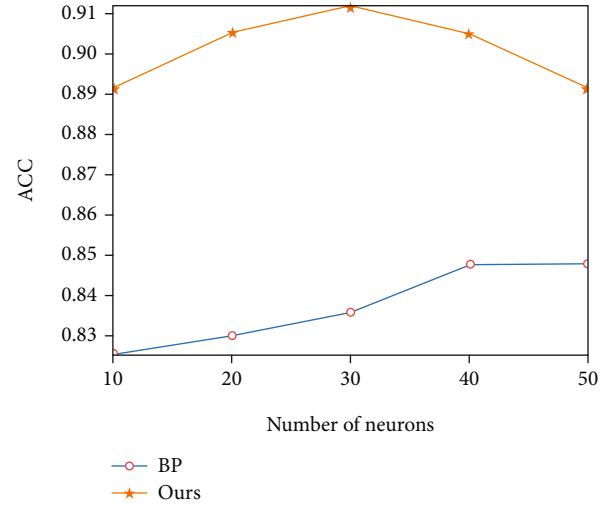


FIGURE 4: Change curve of neurons.

TABLE 2: Results of ablation experiments.

Method	MSE	ACC
Embed MLP	49.51	0.8317
Ours	32.14	0.9120

## 5. Conclusion

In this paper, a neural network model is used to solve complex nonlinear problems in the assessment of college teaching quality. It also performs an in-depth study on evaluating the quality of college instruction. The development of a new adaptive BP neural network model has been completed. To increase the model's convergence speed, optimize the network structure, and assure its stability, the model incorporates the adaptive vector and momentum of the modified gradient descent BP neural network. Then, to guarantee that the model completely evaluates educational activities, the new evaluation index was added to the traditional evaluation index. We also proved the effectiveness and superiority of the proposed model through experiments. In addition, the research in this paper is based on extensive data research and data cleaning in the early stage. The proposed algorithm will be limited to real-time processing in a complex environment.

## Data Availability

The data used to support the findings of this study are included within the article.

## Conflicts of Interest

All the authors do not have any possible conflicts of interest.

## Acknowledgments

This study was supported by the following projects: (1) Research Project of State Administration of Traditional Chinese Medicine with the Theme of “Extensive Learning, Deep Investigation, & Detailed Implementation” (GY-15, Research & Construction of TCM-Characterized Online General Courses) and (2) Science Popularization Project of Shanghai Municipal Science and Technology Commission (17dz2301000, Construction of TCM Science Popularization System Based on Cultural Inheritance & Health Guidance).

## References

- [1] R. Lozano, M. Y. Merrill, K. Sammalisto, K. Ceulemans, and F. J. Lozano, “Connecting competences and pedagogical approaches for sustainable development in higher education: a literature review and framework proposal,” *Sustainability*, vol. 9, no. 10, p. 1889, 2017.
- [2] E. Abad-Segura and M. D. González-Zamar, “Sustainable economic development in higher education institutions: a global analysis within the SDGs framework,” *Journal of Cleaner Production*, vol. 294, p. 126133, 2021.
- [3] A. Syakur and Y. Sabat, “The effectiveness of cooperative learning (STAD and PBL type) on E-learning sustainable development in higher education,” *Journal of Development Research*, vol. 4, no. 1, pp. 53–61, 2020.
- [4] I. Molderez and E. Fonseca, “The efficacy of real-world experiences and service learning for fostering competences for sustainable development in higher education,” *Journal of Cleaner Production*, vol. 172, pp. 4397–4410, 2018.
- [5] H. A. Patrinos and N. Angrist, *Global Dataset on Education Quality: A Review and Update (2000-2017)*, no. article 8592, 2018 World Bank Policy Research Working Paper, 2018.
- [6] M. Beerkens, “Evidence-based policy and higher education quality assurance: progress, pitfalls and promise,” *European Journal of Higher Education*, vol. 8, no. 3, pp. 272–287, 2018.
- [7] I. Degtjarjova, I. Lapina, and D. Freidenfelds, *Student as Stakeholder: “Voice of Customer” in Higher Education Quality Development*, Sumy State University, 2018.
- [8] W. Zimmerman, B. Altman, B. Simunich, K. Shattuck, and B. Burch, “Evaluating online course quality: a study on implementation of course quality standards,” *Online Learning*, vol. 24, no. 4, pp. 147–163, 2020.
- [9] P. M. Bigatel and S. Edel-Malizia, “Using the “Indicators of Engaged Learning Online” framework to evaluate online course quality,” *TechTrends*, vol. 62, no. 1, pp. 58–70, 2018.
- [10] W. J. Van de Grift, S. Chun, R. Maulana, O. Lee, and M. Helms-Lorenz, “Measuring teaching quality and student engagement in South Korea and the Netherlands,” *School Effectiveness and School Improvement*, vol. 28, no. 3, pp. 337–349, 2017.
- [11] E. Salman, “Teaching quality evaluation: online vs. manually, facts and myths,” *Journal of Information Technology Education: Innovations in Practice*, vol. 16, no. 1, pp. 277–290, 2017.
- [12] M. Bameni Moghadam and S. R. Rafiey, “Evaluation of teaching quality of faculty members in higher education institutions: a case study of faculty of economics at Allameh Tabataba’i University,” *Quarterly of Educational Measurement*, vol. 8, no. 29, pp. 1–22, 2017.
- [13] A. Debroy, A. Ingole, and A. Mudey, “Teachers’ perceptions on student evaluation of teaching as a tool for faculty development and quality assurance in medical education,” *Journal of Education and Health Promotion*, vol. 8, 2019.
- [14] N. Huang, “Analysis and design of university teaching evaluation system based on JSP platform,” *International Journal of Education and Management Engineering*, vol. 7, no. 3, pp. 43–50, 2017.
- [15] M. Gao, W. Cai, and R. Liu, “AGTH-Net: attention-based graph convolution-guided third-order hourglass network for sports video classification,” *Journal of Healthcare Engineering*, vol. 2021, Article ID 8517161, 10 pages, 2021.
- [16] C. Yan, G. Pang, X. Bai et al., “Beyond triplet loss: person re-identification with fine-grained difference-aware pairwise loss,” *IEEE Transactions on Multimedia*, p. 1, 2021.
- [17] J. Zhang, J. Sun, J. Wang, and X. G. Yue, “Visual object tracking based on residual network and cascaded correlation filters,” *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 8, pp. 8427–8440, 2021.
- [18] Y. Tong, L. Yu, S. Li, J. Liu, H. Qin, and W. Li, “Polynomial fitting algorithm based on neural network,” *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 1, pp. 32–39, 2021.
- [19] X. Ning, Y. Wang, W. Tian, L. Liu, and W. Cai, “A biomimetic covering learning method based on principle of homology continuity,” *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 1, pp. 9–16, 2021.
- [20] C. Wang, X. Bai, X. Wang et al., “Self-supervised multiscale adversarial regression network for stereo disparity estimation,” *IEEE Transactions on Cybernetics*, pp. 1–14, 2020.
- [21] X. Bai, C. Yan, H. Yang, L. Bai, J. Zhou, and E. R. Hancock, “Adaptive hash retrieval with kernel based similarity,” *Pattern Recognition*, vol. 75, pp. 136–148, 2018.
- [22] L. Zhang, X. Wang, X. Dong, L. Sun, W. Cai, and X. Ning, “Finger vein image enhancement based on guided tri-Gaussian filters,” *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 1, pp. 17–23, 2021.
- [23] A. R. A. Salih, “Peer evaluation of teaching or “fear” evaluation: in search of compatibility,” *Higher Education Studies*, vol. 3, no. 2, pp. 102–114, 2013.

## Research Article

# Research on the Station Layout Method of Ground-Based Pseudolite Positioning System Based on NSGA-II Algorithm

Li Yang,<sup>1</sup> Kaiyuan Yang<sup>1</sup> ,<sup>2</sup> and Danshi Sun<sup>3</sup>

<sup>1</sup>College of Geography and Environmental Science, Henan University, Kaifeng, 475000 Henan, China

<sup>2</sup>Department of Electrical and Electronic Engineering, The University of Sheffield, Sheffield S10 2TN, UK

<sup>3</sup>School of Geodesy and Geomatics, Wuhan University, Wuhan, Hubei 430000, China

Correspondence should be addressed to Kaiyuan Yang; [sbc-18-8022@sbc.usst.edu.cn](mailto:sbc-18-8022@sbc.usst.edu.cn)

Received 28 June 2021; Revised 16 July 2021; Accepted 27 July 2021; Published 15 August 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Li Yang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Given the problem that the existing method of station distributing the pseudosatellite system cannot ensure both its coverage and position in a situation of signal occlusion, it proposed a new stationary layout method with an elite strategy for a ground-based pseudosatellite positioning system based on the elite strategy of the nondominant genetic rankings (NSGA-II). The geometrical design of the pseudosatellite system is calculated by visual domain analysis and precision factors for the signal coverage age and base station. To optimize the algorithm, the NSGA-II algorithm is used. An earth pseudosatellite positioning system method of stationary distribution is obtained that simultaneously optimizes signal coverage and positioning accuracy. The algorithm is better distributed and has a certain superintendence compared with the traditional genetic algorithm.

## 1. Introduction

Ground-based pseudobase station is a kind of ground device placed on the ground and composed of ground transmitting base station and pseudolite receiver. It can provide users with continuous and highly reliable positioning, navigation, and other services. It has the characteristics of strong anti-interference ability, low cost, and high precision and is widely used in various fields. At present, with the application of ground-based pseudobase station positioning system in more and more scenes, in order to obtain positioning more accurately, relevant experts and scholars have done a lot of research on its coverage and geometric spatial layout. For example, in references [1, 2] based on square root UKF algorithm, the tracking and positioning ability of pseudolites is improved by optimizing the spatial layout of pseudolites. In references [3, 4] based on initial value filtering algorithm and second-order digital filter, GNSS assists pseudolite system positioning system to improve system signal coverage. Although the above methods have improved the signal coverage or positioning accuracy of the ground-based pseudolite system to a certain extent, on the whole, the above methods

have only been optimized from a single aspect, and the research on the signal coverage and geometric layout of the ground-based pseudolite system is not comprehensive enough. Based on this, this study proposes a station deployment method for the ground-based pseudolite system based on the NSGA-II algorithm. Taking the coverage function and the accuracy factor as the objective functions [5–7], the NSGA-II algorithm is used for optimization, and the system signal coverage and positioning accuracy can be improved simultaneously [8–10]. In the station deployment method, experiments show that this method can effectively improve the signal coverage and positioning accuracy of the system and provide a new idea for the station deployment of ground-based pseudolite system in various scenes. The NSGA-II algorithm proposed in this paper has good research significance for base station location and signal accuracy and solves the problem of base station location and positioning. The second part of this paper introduces the basic content of the NSGA-II algorithm, the third part introduces the method of ground-based pseudolite positioning system station deployment, and the fourth part compares the application of the NSGA-II algorithm with the other three algorithms.

## 2. NSGA-II Algorithm

Recently, machine learning methods have been developed in many fields, such as image processing [11, 12], remote sensing [13, 14], intelligent transportation [15], and other applications [16–18]. As a branch of machine learning, genetic algorithm has been used in many applications. And NSGA-II algorithm is a multiobjective optimization genetic algorithm, which is often used to solve multiobjective optimization problems. The basic idea is to obtain the optimal individual by simulating natural selection and evolution. The specific steps are as follows:

*Step 1.* Initialize the population. The initial population with scale  $N$  is randomly generated, and the first generation of offspring population is obtained by non-dominated sorting, crossover and mutation operations.

*Step 2.* Merging the parent population and the offspring population and calculating the objective function value of each individual.

*Step 3.* After performing nondominant sorting on the merged population individuals, the crowding distance by Wang [19] is calculated by adopting the following formula:

$$L[i] = \sum_{k=1}^n |f_k(i+1) - f_k(i-1)|, \quad (1)$$

where  $i$  represents an individual of the population;  $i+1$  and  $i-1$  denote the adjacent individuals of  $i$ ; and  $f_k(\cdot)$  denotes the objective function.

*Step 4.* According to the crowded distance, the first  $N$  individuals are reserved to form a new parent population.

*Step 5.* Judging whether the algorithm meets the termination condition, if the maximum iteration times are reached, selecting the optimal solution for output according to the utility function, and ending the algorithm.

The above flow can be illustrated in Figure 1.

The number of iterations is set according to experience or the difference between before and after iterations is less than the threshold to determine whether the number of iterations is met. Generally, setting a larger number of iterations is to achieve a better optimization effect.

## 3. Station Arrangement Method of Ground-Based Pseudolite Positioning System Based on NSGA-II Algorithm

*3.1. Multiobjective Optimization Model of Ground-Based Pseudolite Base Station Layout.* The layout scheme of ground-based pseudolite base stations is usually determined by the system signal coverage rate and the geometric layout of base stations. Therefore, the research is aimed at the most slipper coverage rate and the optimization of base station

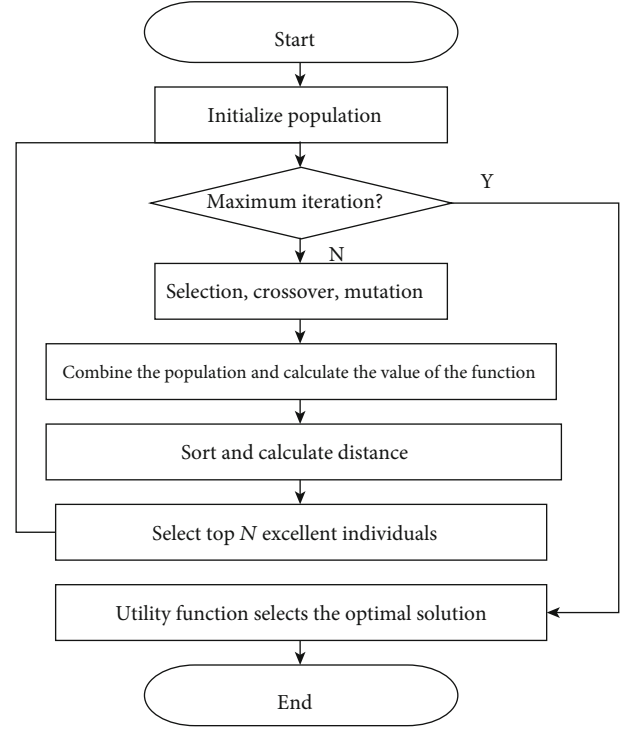


FIGURE 1: NSGA-II algorithm flow.

layout. If the coordinate of the  $i$  pseudolite base station is  $S_i$  and there are  $N$  base stations, the mathematical model of ground-based pseudolite layout can be expressed as [20]

$$\begin{cases} \text{Maximize } f_1(S_i) = \frac{\text{area}(V(S_1, S_2, \dots, S_n))}{\text{area}}, \\ \text{Maximize } f_2(S_i) = g(\text{DOP}(S_1, S_2, \dots, S_n)), \end{cases} \quad (2)$$

where  $f_1$  represents the system signal coverage rate,  $\text{area}(V(S_1, S_2, \dots, S_n))$  reflects the system signal coverage rate and indicates the system signal coverage area;  $f_2$  represents the accuracy factor, reflecting the advantages and disadvantages of the system base station layout; and  $g(\text{DOP}(S_1, S_2, \dots, S_n))$  indicates the positioning accuracy.

### 3.2. Objective Function Design

*3.2.1. Coverage Function Design.* The signal coverage rate of pseudolite system is the ratio of the locatable area of pseudolite system to the target coverage area, which can be solved by visual domain analysis based on digital elevation model (DEM).

Firstly, the locatable area of pseudolite system is determined. DEM data of longitude, latitude, and elevation terrain of the target area are imported and converted into a two-dimensional plane coordinate system relationship, as shown in Equations (3)–(5) [21].

$$\text{DEM}x = [Dx_{ij}] M \times N, \quad (3)$$

$$\text{DEM}y = [Dy_{ij}] M \times N, \quad (4)$$

$$\text{DEM}z = [Dz_{ij}]M \times N, \quad (5)$$

where  $M$  and  $N$  represent the two-dimensional size of the DEM data matrix;  $Dx_{ij}$ ,  $Dy_{ij}$ , and  $Dz_{ij}$  represent the horizontal, ordinate, and height of the terrain raster data, respectively.

Then, visual domain analysis is used to solve the signal coverage of each base station. Let the coordinates of the actual terrain base station  $i$  be  $s_i(x_i, y_i, z_i)$ , as follows:

$$v_i = [a_{ij}]_{M \times N}, \quad (6)$$

where  $v_i$  represents the visual field matrix of the base station  $i$ ; when  $a_{ij} = 1$  or  $a_{ij} = 0$ ,  $a_{ij} = 1$  indicates that the base station is visible to the points ( $Dx_{ij}$ ,  $Dy_{ij}$ , and  $Dz_i$ ), and when  $a_{ij} = 0$ , it indicates that it is invisible; that is, the more the zeros in  $v_i$ , the smaller the visible range and the smaller the signal coverage range.

Finally, the coverage of all base stations is superimposed, as shown in the following equation:

$$v = \sum_{i=1}^n v_i = [v_{ij}]_{M \times N}, \quad (7)$$

where  $v_{ij} \geq [0, N]$  denotes the number of visual pseudolite base stations ( $Dx_{ij}$ ,  $Dy_{ij}$ , and  $Dz_i$ ), where  $N$  denotes the number of base stations. When  $v_{ij} \geq 4$ , it means that the receiver can locate [22]. Therefore, for the convenience of processing, if the value of  $v_{ij} \leq 4$  is set to 1 and the value of  $v_{ij} < 4$  is set to 0, then the matrix  $V$  is  $\bar{V}$ . The set of all 1s in  $\bar{V}$  is the locatable area of the pseudolite system. Therefore, the coverage rate can be calculated by the following equation:

$$\text{Area} = \frac{\text{sum}(\bar{V})}{M \times N} \times 100\%, \quad (8)$$

where sum represents summation. For the convenience of subsequent description,  $f_1$  is transformed into a minimization objective function and defined as

$$f_1 = \frac{1}{\text{area}} = \frac{M \times N}{\text{sum}(\bar{V})} \times 100\%. \quad (9)$$

**3.2.2. Design of Accuracy Factor Function.** Assuming that the number of sampling points is  $N_p$ , the weight corresponding to each sampling point is  $W_i$ , and the precision factor (DOP) value is  $\text{DOP}_i$ , the objective function  $f_2$  can be defined as

$$f_2 = \frac{\sum_i^{N_p} w_i \times \text{DOP}_i}{\sum_i^{N_p} w_i}. \quad (10)$$

It can be seen from the formula that the smaller the  $f_2$  value, the higher the positioning accuracy in the pseudolite system area. Considering the influence of other factors on positioning accuracy, a global weighted DOP method

based on multimatrix multiplication is proposed, as shown in Equation (11) [23]:

$$\begin{cases} M = \prod_i m_i = [M_{ij}]_{M \times N}, \\ \text{DOP} = D(M), \\ f_2 = \frac{\text{sum}(\text{DOP} \cdot M)}{\text{sum}(M)}, \end{cases} \quad (11)$$

where  $m_i$  represents a condition matrix and  $m$  represents a dot multiplication matrix of a plurality of  $m_i$ ;  $D(M)$  denote whether  $\text{DOP}_{ij}$  is calculated according to  $M_{ij}$ . If  $M_{ij} = 0$ ,  $\text{DOP}_{ij}$  is not calculated; let  $\text{DOP}_{ij} = 0$ ; if  $M_{ij} \neq 0$ , then  $\text{DOP}_{ij}$  is calculated. Set  $m_i$  to

$$m_1 = \bar{V},$$

$$m_2 = [1]_{M \times N},$$

$$m_3 = \begin{bmatrix} 1 & 0 & \cdots & 1 & 0 \\ 0 & 1 & & 0 & 0 \\ & \vdots & \ddots & \vdots & \\ 1 & 0 & \cdots & 1 & 0 \\ 0 & 1 & & 0 & 1 \end{bmatrix}, \quad (12)$$

where  $M1$  represents the visual domain matrix of the current pseudolite system, which can prevent unlocatable points from being included in weighted DOP calculation;  $M2$  represents the weight matrix, where all elements are 1; and  $M3$  represents the sampling matrix, in which 1 represents the sampling point position and 0 represents the nonsampling point.

**3.3. Design of Station Arrangement Method for Ground-Based Pseudolite Positioning System Based on NSGA-II.** The purpose of the station arrangement method of the ground-based pseudolite positioning system based on the NSGA-II algorithm is to solve the coordinates of pseudolite base stations, so the station arrangement design can be carried out according to the following steps:

*Step 1.* Population initialization. The coordinates of all pseudolite base stations are sequentially added into a vector to form an individual chromosome vector  $C$ :

$$C = (x_1, y_1, x_2, y_2, \dots, x_n, y_n), \quad (13)$$

where  $x_i$  and  $y_i$  are the abscissa and ordinate of the base station, respectively, and their ranges are determined by DEM data. According to the above formula, the initial population  $Q_0$  with the number of individuals  $N_v$  can be obtained.

*Step 2.* Select  $N_v$  pairs of parent from a parent population  $Q_k$  according to a tournament mechanism and crossing each

pair of parents in a multipoint crossing mode to generate a subpopulation  $R_k$  [24] of  $N_v$  offspring individuals.

*Step 3.* Randomly select individuals from  $R_k$  according to the mutation probability to perform multipoint mutation.

*Step 4.* Combine  $Q_k$  and  $R_k$  into a population  $U_k = 2N_v$ , and determine a nondominant grade of each individual in that  $U_k$  population according to objective functions  $f_1$  and  $f_2$ .

*Step 5.* Calculate the crowded distance of individual in the crowded  $U_k$  population, performing descending order sorting according to the nondominant grade and retaining the first  $N_v$  individuals.

*Step 6.* Calculate utility function values  $F$  of all noninferior solutions by adopting utility functions, such as Equation (14), and select the solution with the largest  $F$  value as the best site selection scheme for pseudolite base stations.

$$F = C_1 \times \frac{f_1(X_i) - f_{1 \min}}{f_{1 \max} - f_{1 \min}} + C_2 \times \frac{f_{2 \max} - f_2(X_i)}{f_{2 \max} - f_{2 \min}}, \quad (14)$$

where  $C_1$  and  $C_2$  represent coefficients,  $C_1 + C_2 = 1$  and  $C_1, C_2 \in [0, 1]$ , which can be determined by preference. If  $C_1 > C_2$ , it means that the base station layout scheme is more inclined to the base station layout scheme with high signal coverage; if  $C_1 < C_2$ , it means that the base station layout scheme is more inclined to the base station layout scheme with high positioning accuracy. If  $C_1 = C_2 = 0.5$ , the coverage and base station layout are relatively balanced, which is beneficial to the coverage and base station layout.

## 4. Simulation Test

*4.1. Experimental Scheme.* In order to verify the application effect of the proposed algorithm in practical application, this experiment takes DEM data of about 15.5 km \* 15.5 km in a mountainous area of Hunan as an example and adopts four strategies to deploy pseudolite base stations. The specific strategies are as follows:

*Strategy 1.* Nine, 16, 25, 36, and 49 pseudolite base stations are evenly placed in the target area as a reference.

*Strategy 2.* Standard genetic algorithm is used to solve the station layout strategy, and only the system coverage rate is optimized.

*Strategy 3.* The standard genetic algorithm is used to solve the station layout strategy, and only the system base station set layout is optimized.

*Strategy 4.* This study uses the NSGA-II algorithm to solve the station deployment strategy and optimizes the system signal coverage and positioning accuracy at the same time.

*4.2. Parameter Setting.* The parameters of standard genetic algorithm and NSGA-II algorithm in strategies 2, 3, and 4

TABLE 1: Policy 1 base station coverage and HDOP relationship.

Number of base stations	Coverage (%)	Weighted average HDOP
9	8.2	5.5
16	14.3	5.1
25	23.4	4.3
36	36.8	3.1
49	52.2	2.8

TABLE 2: Policy 2 base station coverage and HDOP relationship.

Number of base stations	Coverage (%)	Weighted average HDOP
10	52.5	32.3
15	69.4	18.6
20	81.1	27.2
25	89.3	18.1
30	94.1	12.6
35	96.3	16.5
40	97.8	11.8

TABLE 3: Policy 3 base station coverage and HDOP relationship.

Number of base stations	Coverage (%)	Weighted average HDOP
10	44.3	3.6
15	52.5	3.2
20	69.1	2.9
25	73.4	2.5
30	81.5	2.2
35	87.7	1.9
40	89.1	1.45

are set as follows: the number of populations is 80, the genetic algebra is 100, the crossover probability is 0.8, and the mutation probability is 0.01.

### 4.3. Case Analysis

*4.3.1. Algorithm Results.* Using the above four strategies for simulation, the system coverage and weighted average HDOP results are shown in Tables 1–4. As can be seen from the table, the coverage rate of strategy 1 is poor; even if the number of base stations reaches 49, its highest coverage rate is only 52.2%, but its HDOP performs well due to its uniform distribution of base stations. Strategy 2 optimizes the coverage rate, so under the same number of base stations, its coverage rate is the largest compared with other strategies. Strategy 3 optimizes the layout of base station sets, so under the same number of base stations, its HDOP is the smallest and the positioning accuracy of the system is the best. Strategy 4 optimizes the coverage and positioning accuracy of the system at the same time, so compared with other strategies, this strategy has the highest comprehensive level of coverage and positioning accuracy.



TABLE 4: Policy 4 base station coverage and HDOP relationship.

Number of base stations	Coverage (%)	Weighted average HDOP
10	50.5	8.4
15	66.4	6.1
20	76.1	4.5
25	85.3	4.1
30	92.5	3.8
35	94.8	3.4
40	96.8	3.2

As can be seen from Tables 1–4, the coverage rate is directly proportional to the number of base stations, while the weighted average HDOP is counterexample to the number of base stations. Therefore, the coverage rate and weighted average HDOP are inversely proportional. Therefore, the smaller the accuracy and the greater the coverage, the better the performance of the NSGA-II algorithm.

**4.3.2. Comparison of Algorithms.** In order to verify the superiority of the proposed algorithm, the optimization effects of this algorithm and standard genetic algorithm are compared; that is, the optimization effects of strategy 2 and strategy 4 and strategy 3 and strategy 4 are compared.

- (1) Comparison of optimization results between strategy 2 and strategy 4

The optimization results of strategy 2 and strategy 4 are compared, and the results are shown in Figure 2. As can be seen from the figure, the coverage rate of strategy 2 is relatively close to that of strategy 4, and the average coverage rate difference between the two is only 2.69%, indicating that the coverage rate of the two is basically the same. The HDOP of the two strategies is quite different. The HDOP of strategy 2 is obviously higher than that of strategy 4, which is about 4 times that of strategy 4, indicating that its positioning accuracy is low. With the increase of the number of base stations, the HDOP curve of strategy 2 fluctuates greatly, while the HDOP curve of strategy 4 decreases gently, which indicates that the optimization of coverage rate by standard genetic algorithm is divergent. On the whole, the coverage rate of strategy 2 is close to that of strategy 4, but the positioning accuracy of strategy 4 is higher, which indicates that the station deployment method of strategy 4 is better; that is, the station deployment method of the pseudolite positioning system based on the NSGA-II algorithm proposed in this study is better.

- (2) Comparison of optimization results between strategy 3 and strategy 4

The optimization results of strategy 3 and strategy 4 are compared, and the results are shown in Figure 3. As can be seen from the figure, the coverage rate of strategy 3 and strategy 4 is quite different, and the average coverage rate of strategy 3 is reduced by 10, 68% compared with strategy 4. The HDOP of strategy 4 is about 1.8 times that of strategy 3; that

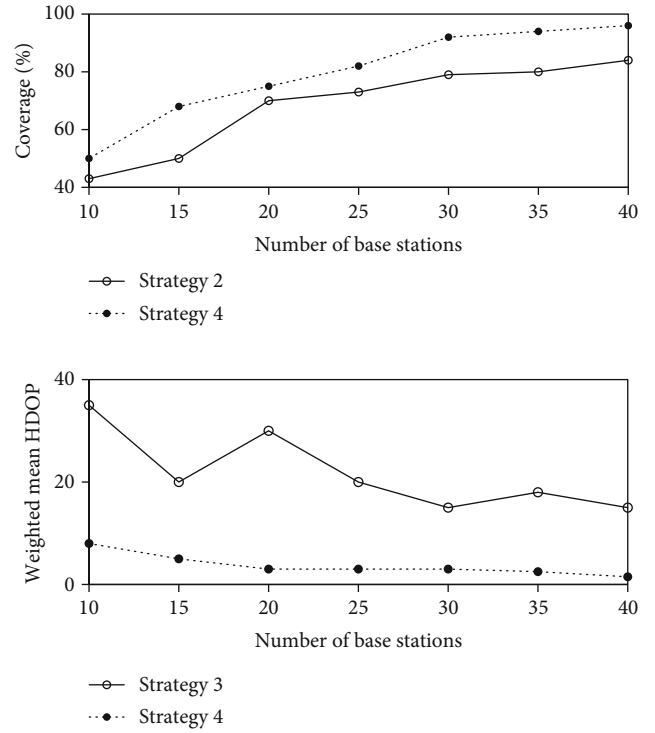


FIGURE 2: Comparison of optimization results of strategy 2 and strategy 4.

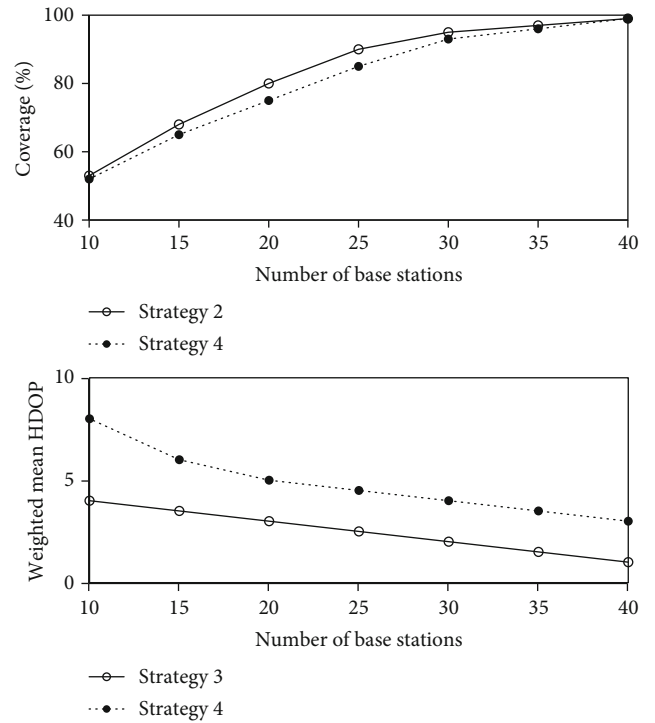


FIGURE 3: Comparison of optimization results of strategy 3 and strategy 4.

is, the difference of positioning accuracy is small. Therefore, although the positioning accuracy of strategy 4 is slightly lower than that of strategy 3, its coverage rate is higher, which

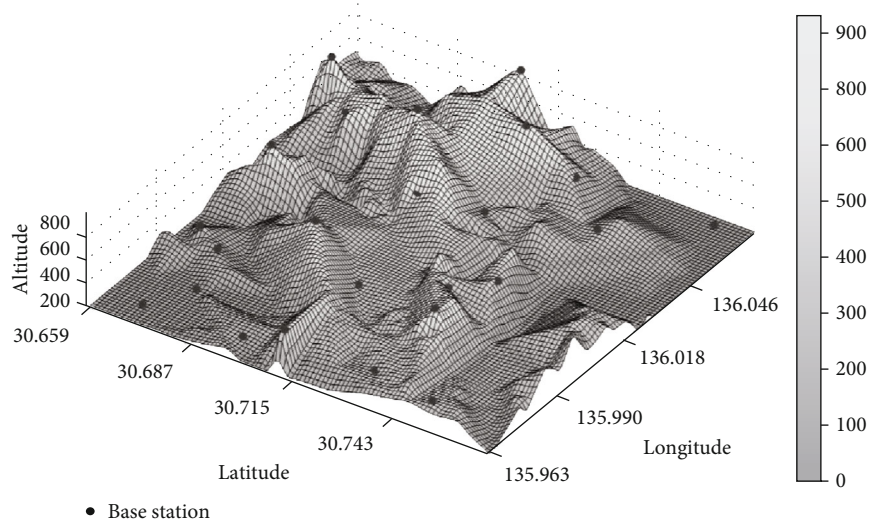


FIGURE 4: Schematic diagram of base station location when the number of base stations = 25.

shows that the station deployment method under this strategy is better; that is, the pseudolite positioning system station deployment method based on the NSGA-II algorithm proposed in this study is better.

Through the above analysis, it can be seen that compared with the station deployment method based on traditional genetic algorithm, the station deployment method of the pseudolite positioning system based on the NSGA-II algorithm in this study can optimize the signal coverage rate and the geometric layout of base stations at the same time, and with the increase of the number of base stations, the signal coverage rate and positioning accuracy are improved more obviously.

**4.4. Algorithm Performance Analysis.** According to the above case analysis results, when the number of base stations is 25, the signal coverage and positioning accuracy can be maximized by using this algorithm. Therefore, in order to verify the performance of the proposed ground-based pseudolite positioning system station deployment method based on the NSGA-II algorithm, the proposed algorithm is used to deploy 25 ground-based pseudolite base stations as an example. When the number of base stations = 25, the location of pseudolite base stations is shown in Figure 4.

According to the simulation results of station distribution, the forefront of Pareto is drawn, as shown in Figure 5. Finally, the individual with  $f_1 = 1.172$  and  $f_2 = 4.1$  in the optimal solution set is selected as the preferred solution by using utility function. As can be seen from the figure, the forefront of this research algorithm is evenly dispersed, indicating that the algorithm has good performance; The range of Pareto solution set objective function  $f_1$  is [1.14, 1.23], indicating that the coverage range of the station deployment scheme is 81.3%~87.7%; The value range of the objective function  $f_2$  is [3.5, 6.3], indicating that the weighted average HDOP range corresponding to the Pareto optimal solution set is [3.5, 6.3].

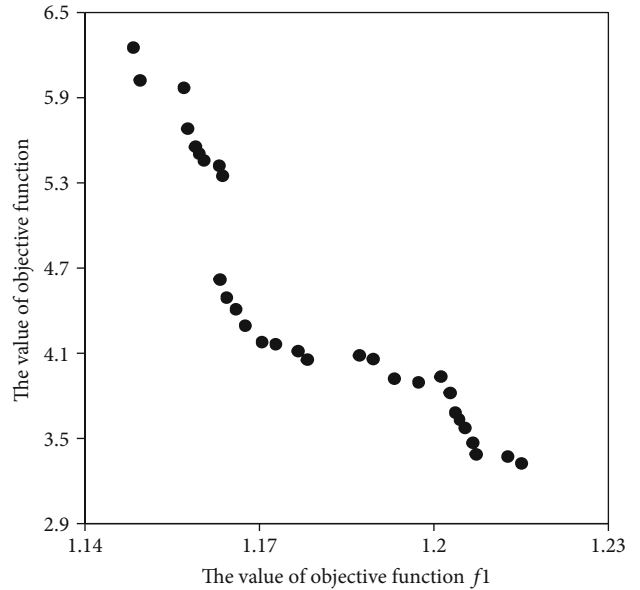


FIGURE 5: Pareto optimal solution set distribution of algorithm when the number of ground-based pseudolite base stations = 25.

## 5. Conclusion

To sum up, the station deployment method of the ground-based pseudolite positioning system based on the NSGA-II algorithm proposed in this study can improve the signal coverage and positioning accuracy of pseudolite base stations by optimizing the coverage rate and geometric layout of base stations. Compared with the standard genetic algorithm, the coverage rate of this algorithm is higher and the positioning accuracy is better. With the increase of the number of base stations, the superiority of this algorithm is more obvious, and the improvement effect of signal coverage rate and positioning accuracy is better. Although this algorithm has achieved some results, there are still some problems in practical application, such as not considering the influence of

electromagnetic interference and other factors in the actual environment and not including the actual installation difficulty of the base station. Therefore, the next step will be to conduct in-depth research from these aspects in order to find a more accurate and effective method for the deployment of ground-based pseudolite systems.

## Data Availability

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Conflicts of Interest

The authors declared that they have no conflicts of interest regarding this work.

## References

- [1] M. Cai, F. He, and L. Wu, "Application of UKF algorithm for target tracking in DTV-based passive radar," in *2009 2nd International Congress on Image and Signal Processing*, pp. 1–4, Tianjin, China, 2009.
- [2] H. Zhigang, C. Dong, P. Chaofeng, C. Long, and W. Shaohua, "State of charge estimation of power Li-ion batteries using a hybrid estimation algorithm based on UKF," *Electrochimica Acta*, vol. 211, pp. 101–109, 2016.
- [3] T. J. Bouska, J. F. Raquet, and P. S. Maybeck, "The use of optimal smoothing and nonlinear filtering in pseudolite-based positioning systems," in *Proceedings of the 59th Annual Meeting of The Institute of Navigation and CIGTF 22nd Guidance Test Symposium*, pp. 435–443, Albuquerque, NM, 2003.
- [4] C. M. Wang, W. C. Xiao, and I. I. R. Second-order, "Second-order IIR notch filter design and implementation of digital signal processing system," *Applied Mechanics and Materials*, vol. 347–350, pp. 729–732, 2013.
- [5] J. Zhang, Y. Liu, H. Liu, and J. Wang, "Learning local-global multiple correlation filters for robust visual tracking with Kalman filter redetection," *Sensors*, vol. 21, no. 4, p. 1129, 2021.
- [6] D. Zhang, X. Zhang, and H. Qi, "A new location sensing algorithm based on DV-hop and quantum-behaved particle swarm optimization in WSN," *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 2, pp. 1–17, 2021.
- [7] J. Zhang, J. Sun, J. Wang, and X. G. Yue, "Visual object tracking based on residual network and cascaded correlation filters," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–14, 2020, In press.
- [8] Y. Li and J. Cao, "WSN node optimal deployment algorithm based on adaptive binary particle swarm optimization," *ASP Transactions on Internet of Things*, vol. 1, no. 1, pp. 1–8, 2021.
- [9] J. Zhang, W. Wang, C. Lu, J. Wang, and A. K. Sangaiah, "Light-weight deep network for traffic sign classification," *Annals of Telecommunications*, vol. 75, no. 7–8, pp. 369–379, 2020.
- [10] G. Li and P. Zhang, "Design and implementation on patchouli growth environment monitoring system based on IoT," *ASP Transactions on Internet of Things*, vol. 1, no. 1, pp. 14–18, 2021.
- [11] S. Qi, X. Ning, G. Yang et al., "Review of multi-view 3D object recognition methods based on deep learning," *Displays*, vol. 69, p. 102053, 2021.
- [12] Y. Jiang, X. Gu, D. Wu, W. Hang, J. Xue, and S. Qiu, "A novel negative-transfer-resistant fuzzy clustering model with a shared cross-domain transfer latent space and its application to brain CT image segmentation," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 18, no. 1, pp. 40–52, 2021.
- [13] W. Cai, Z. Wei, R. Liu, Y. Zhuang, Y. Wang, and X. Ning, "Remote sensing image recognition based on multi-attention residual fusion networks," *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 1, pp. 1–8, 2021.
- [14] R. Liu, X. Ning, W. Cai, and G. Li, "Multiscale dense cross-attention mechanism with covariance pooling for hyperspectral image scene classification," *Mobile Information Systems*, vol. 2021, 15 pages, 2021.
- [15] J. Chen, C. Du, Y. Zhang, P. Han, and W. Wei, "A clustering-based coverage path planning method for autonomous heterogeneous UAVs," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–11, 2021, In press.
- [16] Z. Zhongxiang, F. Chenghua, Z. Liang, and M. Kong, "K algorithm for microstrip peocessor design," *ASP Transactions on Neural Information Computing*, vol. 1, no. 1, pp. 15–21, 2021.
- [17] M. Gao, W. Cai, and R. Liu, "AGTH-Net: attention-based graph convolution-guided third-order hourglass network for sports video classification," *Journal of Healthcare Engineering*, vol. 2021, 10 pages, 2021.
- [18] L. Li, "DSP for condition monitoring equipment in real-time," *ASP Transactions on Neural Information Computing*, vol. 1, no. 1, pp. 22–25, 2021.
- [19] T. Wang, Z. Yao, and M. Lu, "Mesh topology based clock synchronization technique for pseudolite systems," *Navigation*, vol. 67, no. 3, pp. 619–632, 2020.
- [20] S. Ge and Y. Piao, "Positioning technology research of Mobile vehicle group," *Applied Mechanics and Materials*, vol. 716–717, pp. 1451–1454, 2014.
- [21] H. Ding and X. Gu, "Improved particle swarm optimization algorithm based novel encoding and decoding schemes for flexible job shop scheduling problem," *Computers & Operations Research*, vol. 121, p. 104951, 2020.
- [22] B. Situ and W. Jin, "Optimization model for the vehicle scheduling problem in public transportation network," in *2009 International Conference on Management and Service Science*, pp. 1–4, Beijing, China, 2009.
- [23] U. Ahmed, J. C. Lin, J. M. Wu, Y. Djenouri, G. Srivastava, and S. K. Mukhiya, "Efficient mining of Pareto-front high expected utility patterns," in *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, pp. 872–883, Cham, 2020.
- [24] C. Q. Gong, Z. C. Jiang, C. H. Huang, P. Wang, and T. Wu, "Optimization analysis of GDOP of PL-aided navigation and positioning system," *Applied Mechanics & Materials*, vol. 239, pp. 544–547, 2013.

## Research Article

# Sentiment Analysis of Chinese Paintings Based on Lightweight Convolutional Neural Network

Jiaying Bian  and Xiaoying Shen 

Wuxi Vocational College of Science and Technology, No. 8 Xinxu Road, Wuxi, Jiangsu 214000, China

Correspondence should be addressed to Xiaoying Shen; 1201801@wxsc.edu.cn

Received 13 July 2021; Revised 28 July 2021; Accepted 30 July 2021; Published 10 August 2021

Academic Editor: Shan Zhong

Copyright © 2021 Jiaying Bian and Xiaoying Shen. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Chinese painting is one of the representatives of our country's outstanding traditional culture, and it embodies the long history and intellectual wisdom of the Chinese nation. In the paper, we combine the artistic characteristics of Chinese paintings and use an optimized SqueezeNet model to study the sentiment analysis of Chinese paintings. To make full use of the advantages of lightweight convolutional neural networks, we make two optimizations based on SqueezeNet. On the one hand, expand the model width to obtain more effective Chinese painting sentiment features for classification tasks, thereby improving the classification accuracy of the model. On the other hand, introduce the idea of residual network to prevent gradient disappearance and gradient explosion in the training process, thereby enhancing the model's generalization ability. To verify the effectiveness of the optimized SqueezeNet model used in the sentiment analysis of Chinese paintings, four kinds of sentiment classifications were carried out on the multitheme Chinese paintings downloaded on the Internet. The results of comparative experiments show that the optimized SqueezeNet model used in this paper can improve the accuracy of classification and has better generalization ability. Finally, the research results of this paper can be applied to the protection of traditional culture, the appreciation of traditional Chinese painting, and art education and training, which is conducive to the inheritance and innovation of the national quintessence and promotes the prosperity and development of traditional art and culture.

## 1. Introduction

Traditional Chinese painting is the artistic treasure of our country. It not only vividly depicts the long history of our country with superb artistic skills but also embodies the ideology, philosophy, cultural concept, and aesthetic characteristics of the Chinese nation. It can be said that Chinese painting is an irreplaceable and important part of the traditional cultural and spiritual heritage of the Chinese nation and stands in the gallery of paintings in the world with its unique and distinctive artistic style. With the rapid development of technology, a large number of digital Chinese paintings have appeared in the Internet and digital museums. However, how to efficiently use and manage these paintings and then to promote the charm of Chinese paintings has become an urgent problem to be solved. Studying the algorithms based on sentiment analysis of Chinese paintings will not only help improve users' ability to learn and appreciate

Chinese paintings but also help promote the construction of digital museums and cultural relic management. It also helps to demonstrate cultural self-confidence and promote the implementation of the cultural power strategy, thereby comprehensively enhancing the soft power and international influence of our country's culture.

Different from Western painting's "write shape by shape," traditional Chinese painting emphasizes imagery. The author often creatively interprets the object artistically in the paintings. It can be said that Chinese painting has been an art form centered on emotion since ancient times. Excellent Chinese painting works are the spiritual product of the artist's depiction with true feelings. The painter does not simply describe the objective things but more importantly expresses his inner feelings, so as to achieve a state of combining things with himself and sublimating the spirit. Because of the artistic characteristics of traditional Chinese paintings, which emphasize "freehand brushwork" and

“focusing on emotional expression,” we inevitably need to consider the emotional factors in the painting when we appreciate, research, and protect Chinese paintings. Understanding high-level emotional semantics can help us better appreciate and study Chinese paintings. Meanwhile, it can also enrich the algorithms on organization, management, and retrieval of Chinese paintings and popularize the dissemination of Chinese paintings.

In view of this, in order to fully understand the emotions contained in Chinese paintings, scholars have done a lot of research work and achieved fruitful research results. Literatures [1, 2] extract local and global features based on histograms to characterize different aspects of the artistic style of Chinese paintings and use these features to drive neural networks to classify ink paintings. Literature [3] proposed an art descriptor based on the characteristics of composition and painting objects, dealing with the correlation and synergy among all the elements in the integrated features, and then combined the Monte Carlo convex hull feature selection model to classify the authors of Chinese paintings. Literature [4] used the STASM algorithm to extract the feature points of the face and the face in Chinese paintings and then converted the style of the user’s facial photo and integrated it into the Chinese painting template to form the user’s ink portrait. Literature [5] used a deep learning model to classify the authors of traditional Chinese ink paintings. They took the stroke image of each input painting as the recognition basis and then input all the strokes into the CNN-based feature extractor to form the combined high-dimensional features of each Chinese painting. Literature [6] used a hyperspectral camera with a specific frequency of visible light to scan Chinese paintings and combined the principal component analysis algorithm and the spectral and spatial features extracted by CNN to identify the authenticity of Chinese paintings. Literature [7] introduced the bottleneck layer idea of the Inception module in GoogLeNet on the basis of the deep convolutional encoder-decoder, in order to achieve the purpose of reducing model parameters and speeding up the calculation. Literature [8] proposed an algorithm to simulate the creative process of color ink painting. This algorithm uses CNN and Generative Adversarial Networks (GAN) to convert line and color styles and stylizes flower photos into color ink paintings. Literature [9] used Sobel edge detection to obtain the information of Chinese brush painting and then transformed the information by discrete cosine transform as the input of CNN and finally combined support vector machine to classify the two styles of Chinese painting meticulous and freehand. This algorithm establishes a hybrid model composed of CNN and support vector machine and only uses the stroke features in the painting. Literature [10] proposed an algorithm of joint mutual information and data-embedded classification. They first used the VGG model to extract the features of Chinese paintings and then used mutual information theory to make the distribution information of the image affected by the importance of the features, thereby improving the classification accuracy. Literature [11] proposed a style transfer algorithm for Chinese paintings. It mainly inputs the four restriction conditions of ink painting strokes, white space, ink smearing, and yellow-

ish tone into CNN, designs corresponding conversion strategies for different styles of freehand brushwork and meticulous brushwork, and obtains better ink painting visual effects.

The above research shows that the deep network structure can obtain more abstract high-level semantic features; it can handle more complex classification tasks. Therefore, deep convolutional network has obvious advantages in tasks such as feature extraction and classification. In order to improve the classification efficiency of the CNN in the application of Chinese painting sentiment classification, we use an optimized lightweight CNN to recognize the sentiment of Chinese paintings. The related research work is as follows:

- (1) Firstly, the theory of image sentiment analysis and related algorithms are studied to provide a theoretical basis for follow-up research
- (2) Secondly, combining the artistic characteristics of Chinese paintings, try to use a lightweight CNN to study the sentiment classification of Chinese paintings. At the same time, the researchers also made two optimizations based on SqueezeNet. On the one hand, expand the model width to obtain more effective Chinese painting sentiment features for classification tasks, thereby improving the classification accuracy of the model. On the other hand, introduce the idea of residual network to prevent gradient disappearance and gradient explosion in the training process, thereby enhancing the model’s generalization ability
- (3) Finally, the optimized lightweight CNN is applied to the sentiment recognition of Chinese paintings. The experimental results show that compared with other deep learning algorithms, the classification accuracy and efficiency of the algorithm used in this paper are relatively high in the task of Chinese painting sentiment classification, verifying the feasibility and effectiveness of the algorithm

## 2. Related Work

*2.1. Image Sentiment Analysis.* At present, most sentiment analysis tasks are based on text content, and many scholars have conducted comprehensive and in-depth research in the field of text sentiment analysis. However, as a special image classification technology, image sentiment analysis is still under constant exploration in both technical and application aspects. Different from text sentiment analysis, the emotion in the picture is often hidden, which is a kind of high-level semantic emotional understanding. Studying the sentiment of massive images can not only understand the emotional needs of users but also help optimize user experience in image retrieval and recommendation systems. What is more, it is also helpful for studying hot social issues, learning about the attitudes of the public, and providing data support for online public opinion analysis and monitoring.

Image sentiment analysis is a multidisciplinary task. Since the understanding and classification of emotions in pictures are affected by subjective, objective, and cultural

factors in aesthetics, image sentiment analysis involves many disciplines such as artificial intelligence, computer vision, psychology, and aesthetics. Image sentiment analysis mainly refers to the fact that the computer recognizes the content expressed in digital images to cause people's emotional responses, and then, according to different emotional responses, the images can be classified to different classes. Generally, there is a big difference between the feature information extracted by the computer from the low-level visual data and the human emotion interpretation [12, 13]. We call the difference the emotional gap [14, 15]. We hope that the computer can understand the image in depth and accurately recognize the emotional semantics contained in the image. At the same time, on the basis of understanding the emotional semantics of the image, the sentiment classification of the image is carried out according to the different understanding of the emotional semantics of the digital image. In other words, image sentiment analysis is a high-level semantic understanding of the information conveyed by the image. It can bridge the huge emotional gap between low-level visual features and high-level emotions. Today, when the amount of information is increasing rapidly, image sentiment classification is helpful for image labeling and retrieval, which has great social and commercial value and has attracted widespread attention [16].

### 2.2. The Related Methods Based on Image Sentiment Analysis.

Image sentiment analysis methods mainly include methods based on traditional machine learning [17, 18] and methods based on deep learning [19, 20]. Traditional image sentiment analysis methods first need to extract different dimensional visual features from digital images and use machine learning algorithms and models to train the emotions implicit in the digital images through the mapping of the sentiment space model. For example, literature [21] constructed a color matching emotional word data set based on the relationship between color and emotion and then found the corresponding emotional word in the data set to predict the sentiment of the painting by creating a color spectrum for Western paintings. Literature [22] extracted facial expression action unit features and found the relationship between them to realize video sentiment recognition. Literature [23] used directed acyclic graph support vector machine for facial sentiment classification of connected features. The algorithm extracts facial expressions based on geometric and texture features and proves that simple feature stitching can significantly improve the efficiency of facial sentiment classification. However, traditional image sentiment analysis methods do not take into account the gap between the low-level visual features and the high-level emotional semantics. The extraction of visual features is an important prerequisite for effective sentiment analysis. There is no clear limit on the degree of influence of visual features of different dimensions on the emotion of an image.

With the continuous expansion of the application field of deep learning, researchers have tried to use deep learning algorithms to automatically extract image features and realize sentiment classification, which have achieved better sentiment prediction results. Literature [24] proposed the progressive CNN deep model on the basis of CNN and

constructed a large artificially labeled Twitter visual sentiment data set. Literature [25] integrated the crossresidual neural network into multitask deep learning to solve the classification problem of image objects and their sentiment. Literature [26] analyzed the performance of the CNN layer by layer and fine-tuned the CNN to be applied to image sentiment prediction tasks. It proves the effectiveness of deep network learning to recognize sentiment-related features of natural images. Literature [27] used binary classification to assist in multiclassification tasks of natural image sentiment. However, the data set requires two sets of sentiment labels, binary classification and multiple classification, which increases the workload. Literature [28] combined the visual attention mechanism guided by the saliency map with the CNN architecture to achieve better sentiment classification performance based on natural images. In order to maximize the extraction of features that can represent image emotions, literature [29] proposed a cropping method that uses a fully convolutional network to select emotional regions from an image and uses the interdependence of tags to construct a structured learning model. Aiming at the local image sentiment classification, literature [30] used the feature pyramid network to extract multilayer depth features to remove redundant nonemotional areas. Literature [31] proposed an 11-layer CNN model with visual attention to solve the problem of facial expression recognition. The network model extracts CNN features from face images, calculates the region of interest, and finally, uses the features in the resulting region to determine facial expressions. Literature [32] combined the color histogram and the bottom layer features of local binary pattern texture features with deep sentiment features to identify image sentiment. However, the multilevel feature extraction and fusion steps are scattered and unsystematic. Literature [33] combined the features extracted according to the art theory with CNN features and used support vector machines to recognize image sentiment. However, this algorithm requires manual annotations such as eye movement trajectories, which reduces the practical applicability.

To sum up, the limitation of image sentiment analysis is that the elements that affect sentiment judgment include not only the information of the objects in the image, but the scene of the image can also trigger different emotions. Moreover, different objects in the same image may represent different sentiment classifications, such as the semantic segmentation problem in image sentiment analysis, which needs to be further studied by researchers.

## 3. Sentiment Analysis Model of Chinese Painting Based on Lightweight Convolutional Neural Network

*3.1. Lightweight Convolutional Neural Network.* The lightweight network SqueezeNet was proposed by Stanford and Berkeley researchers in 2017 [34]. The lightweight CNN model redesigns the network structure on the basis of the existing CNN structure to achieve the goal of reducing the amount of parameters and reducing the computational complexity. The core building module of SqueezeNet is Fire

module, which mainly consists of two parts: squeeze layer and expand layer. The specific operations of the two layers are shown in Figure 1. The squeeze layer uses a  $1 \times 1$  convolution kernel to convolve the input features. The main purpose of this is to reduce the number of channels of input features, that is, dimensionality reduction. The expand layer uses  $1 \times 1$  convolution operation and  $3 \times 3$  convolution operation, respectively, and then concatenates the convolution results. The combination of these two layers can effectively reduce the amount of parameters.

As shown in Figure 1, the input feature size of the Fire module is  $H \times W \times M$ , and the output feature size is  $H \times W \times (e_1 + e_3)$ . It can be seen that the spatial size of the feature data before and after is unchanged; only the number of channels, that is, the number of dimensions, is changed. First, suppose that a feature map with a size of  $H \times W \times M$  passes through the squeeze layer; then,  $S_1$  feature maps are obtained. That is, the space size of the feature data remains unchanged, the number of channels changes from  $M$  to  $S_1$ , where  $S_1$  is smaller than  $M$ , so as to achieve the compression effect. Here is a brief introduction to feature map. In each convolutional layer in the CNN, the data exists in three-dimensional form. We can think of it as a stack of many two-dimensional images, each of which is called a feature map. Then, input the feature data of size  $H \times W \times S_1$  into the expand layer, and go through a  $1 \times 1$  convolution operation and a  $3 \times 3$  convolution operation, respectively. Then, the results of the two convolutions are spliced as the output of the Fire module; that is, an output feature of size  $H \times W \times (e_1 + e_3)$  is obtained. The Fire module has three adjustable parameters:  $S_1$ ,  $e_1$ , and  $e_3$ . They, respectively, represent the number of convolution kernels and also represent the dimension of the corresponding output feature map. In SqueezeNet,  $e_1 = e_3 = 4S_1$ .

The core idea of SqueezeNet is to connect multiple Fire modules in a cascaded form. It can give full play to the characteristics of the Fire module and reduce the amount of parameters in the network. The structure of SqueezeNet is shown in Figure 2. SqueezeNet starts with an independent convolutional layer (conv1) and then cascades 8 Fire modules, namely, fire2-fire9. Meanwhile, end the cascade with a convolutional layer (conv10). Finally, the global average pooling layer is used instead of the fully connected layer for output. From the beginning to the end of the network, gradually increase the number of convolution kernels in each Fire module. And after the conv1, fire4, and fire8 layers, the maximum pooling with a step size of 2 is used, respectively, and the global average pooling is performed after conv10. That is, the pooling layer is placed in the back position. The purpose of this is to provide a larger activation map for the convolutional layer. Because the activation map retains a lot of information, the larger the activation map of the convolutional layer, the higher the classification accuracy of the network. It should be noted that the Relu activation function is used in both the squeeze layer and the expand layer of SqueezeNet. After the Fire9 module, the Dropout technology was used, and the output specifications of the Dropout depend on the dimensions of the corresponding convolutional layer.

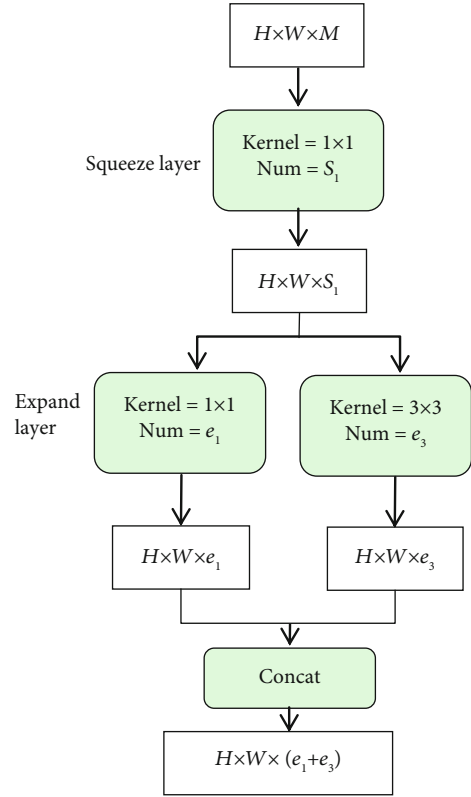


FIGURE 1: Specific operations of squeeze layer and expand layer.

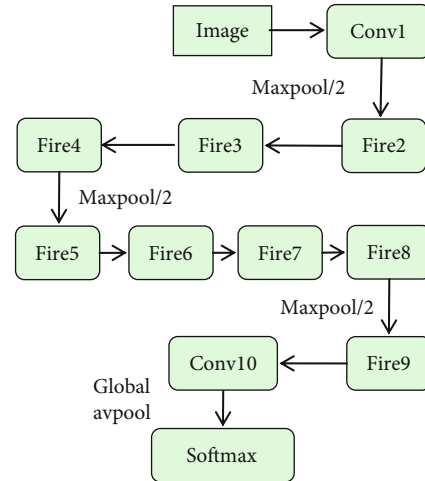


FIGURE 2: The structure of SqueezeNet.

In addition, in order to compress the network SqueezeNet, we also discarded the fully connected layer. Finally, the Softmax classifier is used to output the classification results, and the output specification depends on the classification classes. In the paper, we divide Chinese paintings into four emotions: sad and arrogant, arrogant, lively and cheerful, and quiet and peaceful. That is, the number of classes is 4.

3.2. *ResNet*. The proposal of residual network (ResNet) effectively solves the problem of gradient explosion or

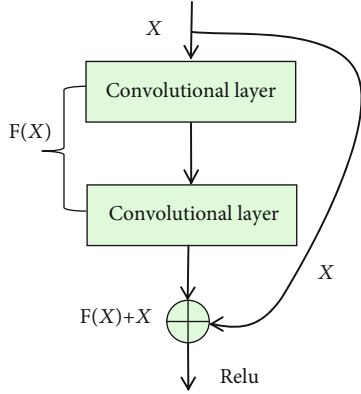


FIGURE 3: Residual error structure.

disappearance in deep network training. Although the deepening of the network structure can improve the accuracy of the network's classification of images, it is not feasible to blindly increase the depth of the network to improve the classification accuracy of the network. On the one hand, as the number of network layers deepens to a certain extent, its performance will not increase linearly with its depth. On the other hand, the deep network will have the problem of gradient disappearance or explosion during the training process. Therefore, deep networks are difficult to train, and ResNet was born to solve this problem. The core idea in ResNet is the residual idea, and the residual structure is shown in Figure 3.

The core idea of the residual structure is to introduce an "identical shortcut connection." For a convolutional layer, the new data  $H(X)$  is obtained after the convolution operation of input data  $X$ . Now, add an identity mapping so that the function  $H(X)$  is transformed into  $F(X) + X$ , where  $F(X)$  is the residual. As the number of network layers increases, if a certain deep network model reaches the optimal level and continues to deepen the number of layers in the network model, even if the residual  $F(X)$  approaches 0, the additional layers will only be equivalent. For the identity mapping operation, at least it will not affect the expressive ability of the model. Moreover, the residual error often does not become 0 during the actual training process, so the convolutional layer can further extract new and more abstract feature information, thereby further improving the representation ability of the model. The proposal of this idea makes it possible to increase the accuracy of the model by increasing the depth of the model.

**3.3. SqueezeNet Optimization Design Based on Sentiment Classification of Traditional Chinese Paintings.** In this section, we mainly optimize the lightweight network SqueezeNet in two parts. First, expand the width of the model. That is, the Fire module is added on the basis of the original network structure to improve the accuracy of model classification. Second, introduce the idea of residual network to prevent gradient disappearance and gradient explosion in the training process, thereby enhancing the generalization ability of the model for image classification and improving the efficiency of the model. The structure of the optimized

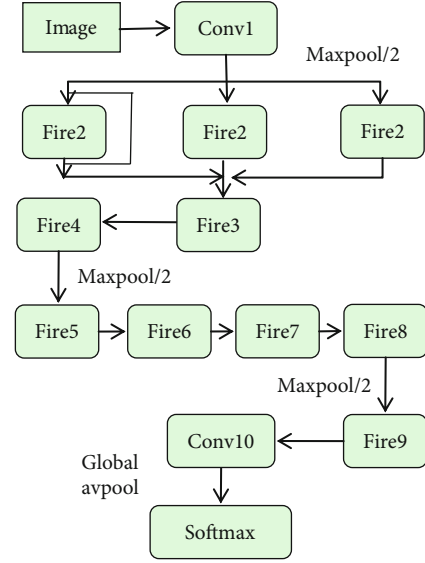


FIGURE 4: The structure of the optimized SqueezeNet based on sentiment classification of traditional Chinese paintings.

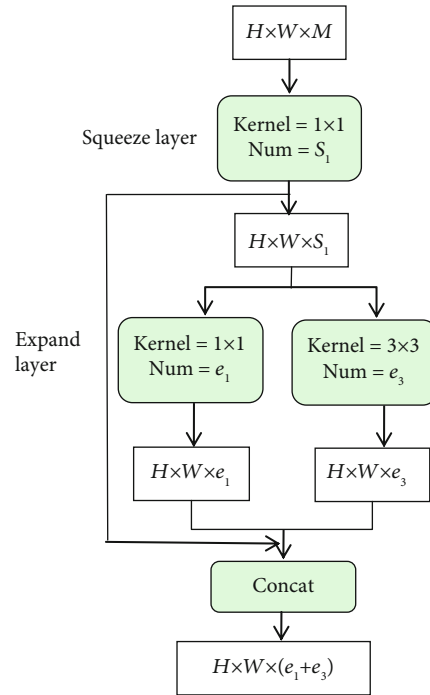


FIGURE 5: The improved Fire module structure.

SqueezeNet based on sentiment classification of traditional Chinese paintings is shown in Figure 4.

As shown in Figure 4, the optimized SqueezeNet did not increase the depth of the network but appropriately increased the width of the network. In addition, the idea of residual network is introduced into the Fire module, and the layer-jumping structure is added to prevent the gradient from disappearing during the training process, which is intended to improve the efficiency of the network. The improved Fire module structure is shown in Figure 5.





FIGURE 6: The representative works of different emotions.

## 4. Experimental Results and Analysis

**4.1. Data Set.** In this paper, we obtained Chinese paintings with multiple themes and periods from the Internet to form a Chinese painting sentiment data set. The data set collected a total of 800 samples of Chinese paintings, including portrait paintings, landscape paintings, flower and bird paintings, and animal paintings. The fine brushwork and freehand painting techniques are all involved. The sample data is extensive, and the quality of images is clear. Based on the analysis of the inscriptions and poems of the paintings, the painter's life experience, and the appreciation and comments of the works, the sample data is divided into four types of common emotions in traditional Chinese paintings: sadness and arrogance, pride and wanton, lively and cheerful, and quiet and peaceful. And the amount of data in each class is equal to ensure the balance of the data set. Figure 6 shows the representative works of different emotions.

**4.2. Experimental Results and Analysis.** In the experiments, we use classification accuracy, network parameters, and network calculations as evaluation indicators to verify the performance of the optimized SqueezeNet in the sentiment analysis of traditional Chinese paintings. Meanwhile, the sentiment class marked with the guidance of the expert is regarded as the correct label. The amount of network calculations refers to the number of floating-point operations required to infer an image. The network calculation amount is obtained by the corresponding relationship among the number of input channels, the number of output channels,

TABLE 1: Classification accuracy of different network models.

Network models	Classification accuracy
AlexNet	0.76
ResNet	0.82
SqueezeNet	0.74
Optimized SqueezeNet	0.86

the height and width of the convolution kernel, and the height and width of the output channels.

First of all, the optimized SqueezeNet and AlexNet, ResNet, and SqueezeNet are, respectively, used to perform classification accuracy experiments on the test set. The classification accuracy of different network models is shown in Table 1.

Analyzing the experimental results in Table 1, it is found that the classification accuracy of the other three models is lower than that of the optimized SqueezeNet on the task of Chinese painting sentiment recognition, since the network model used in the paper is optimized on the basis of SqueezeNet. By expanding the width of the network model and slightly increasing the model parameters, we can obtain more effective Chinese painting sentiment features for classification tasks, thereby improving the classification accuracy of the model. Therefore, the optimized SqueezeNet has a better classification effect and is more suitable for sentiment analysis of Chinese paintings.

Finally, the network parameters and calculations of AlexNet [35], ResNet [36], SqueezeNet [34], and the

TABLE 2: The parameters and calculations of different network models.

Network models	Parameters/ $M$	Calculations/GFLOPI
AlexNet	63.4	1.05
ResNet	13.47	5.24
SqueezeNet	3.63	0.91
Optimized SqueezeNet	3.80	0.72

optimized SqueezeNet are compared, whose results are shown in Table 2.

As shown in Table 2, AlexNet has the largest amount of parameters among the four networks, and the amount of calculations is at the middle level of the four networks. This is because AlexNet contains three fully connected layers, and the parameters in the three fully connected layers account for a relatively large amount, which results in a large amount of overall parameters of AlexNet. The amount of calculations is not very large because the number of layers in AlexNet is small and the parameters of the convolutional layer are small. The amount of parameters of ResNet is at the middle level of the four networks, but the amount of calculations is the largest. This is because the global average pooling layer is applied in ResNet, which effectively reduces the amount of structural parameters. The largest amount of calculation is due to the large number of layers in the ResNet. The amount of parameters in SqueezeNet is very small, and the amount of calculations is among the four networks. This is because the Fire module in SqueezeNet uses less  $3 \times 3$  convolution, so its parameter is very small. The large amount of calculation is because there are more  $1 \times 1$  convolutions in the Fire module. The optimized SqueezeNet introduces the idea of residual network and increases the network width, so the amount of parameters is slightly higher than that of SqueezeNet, but the amount of calculation is the smallest.

In summary, compared with several other network models, the optimized SqueezeNet model used in the paper has obvious advantages in both classification accuracy and recognition efficiency based on the Chinese painting sentiment classification. Therefore, it has certain research significance and practical value.

## 5. Conclusion and Outlook

As one of our country's traditional culture, Chinese painting embodies the long history and culture of our country and carries the thoughts, feelings, and humanistic spirit of the Chinese nation. Emotion can sublimate the artistic value of Chinese painting to a certain extent. Therefore, analyzing the emotions contained in Chinese paintings and assisting in the appreciation of art works are of great significance to the digital management of Chinese paintings and the promotion of the spirit of Chinese paintings. In this paper, combining the artistic characteristics of Chinese paintings and using a lightweight CNN, the sentiment characteristics of Chinese paintings are decomposed and analyzed in a quantified form, and a good sentiment classification result based on Chinese paintings is obtained. On the one hand, it can help the audi-

ence clearly understand the emotions and expressions of Chinese paintings, so as to better appreciate the art of Chinese paintings. On the other hand, it can directly display the regular pattern of Chinese paintings' sentiment, which is helpful for the digital management of Chinese paintings and the protection and dissemination of traditional Chinese culture and fully demonstrates the cultural heritage and spiritual outlook of the nation. However, in the future work, there are still the following issues that need to be further studied and improved.

Firstly, because the characteristics of Chinese paintings in different techniques or themes are very different, it is difficult to recognize their emotions with the same algorithm. In the future, we will study the sentiment recognition algorithms for multiple classes of Chinese paintings and perform more accurate calculations on the sentiment of Chinese paintings, so as to create greater value in the digital management and protection of Chinese paintings.

Secondly, we will further expand the database samples and establish a mathematical model based on the sentiment of Chinese paintings, so as to contribute to the future development of Chinese painting research.

Finally, in future work, virtual reality or augmented reality technology will be combined to improve the experience of visualizing the sentiment of traditional Chinese paintings, so as to promote the digital construction of museum cultural relics and strengthen the social and educational functions of digital museums.

## Data Availability

The labeled data set used to support the findings of this study is available from the corresponding author upon request.

## Conflicts of Interest

The authors declare no conflicts of interest.

## Acknowledgments

This work was supported in part by the National Social Science Fund Project under grant 17BGL102, Excellent Project of Jiangsu Province Social Science Union under grant 15SYC-043, Soft Science Research of Wuxi Science and Technology Association under grant KX15-B-01, and Fundamental Research Funds for the Central Universities under grant 2015ZX18.

## References

- [1] J. Sheng and J. Jiang, "Style-based classification of Chinese ink and wash paintings," *Optical Engineering*, vol. 52, no. 9, article 093101, 2013.
- [2] J. Sheng and J. Jiang, "Recognition of Chinese artists via windowed and entropy balanced fusion in classification of their authored ink and wash paintings (IWPs)," *Pattern Recognition*, vol. 47, no. 2, pp. 612–622, 2014.
- [3] M. Sun, D. Zhang, Z. Wang, J. Ren, and J. S. Jin, "Monte Carlo convex hull model for classification of traditional Chinese paintings," *Neurocomputing*, vol. 171, pp. 788–797, 2016.

- [4] P. Y. Chiang, C. V. Lin, and C. H. Tseng, "Generation of Chinese ink portraits by blending face photographs with Chinese ink paintings," *Journal of Visual Communication and Image Representation*, vol. 52, pp. 33–44, 2018.
- [5] M. Sun, Z. Dong, J. Ren, W. Zheng, and J. S. Jin, "Brushstroke based sparse hybrid convolutional neural networks for author classification of Chinese ink-washpaintings," in *2015 IEEE International Conference on Image Processing (ICIP)*, pp. 626–630, Quebec City, QC, Canada, 2015.
- [6] Z. Wang, D. Lu, D. Zhang, M. Sun, and Y. Zhou, "Fake modern Chinese painting identification based on spectral–spatial feature fusion on hyperspectral image," *Multidimensional Systems and Signal Processing*, vol. 27, no. 4, pp. 1031–1044, 2016.
- [7] R. Zhou, J. H. Han, H. S. Yang, W. Jeong, and Y. S. Moon, "Fast style transfer for Chinese traditional ink painting," in *2019 IEEE 9th International Conference on Electronics Information and Emergency Communication (ICEIEC)*, pp. 586–588, Beijing, China, 2019.
- [8] C. Zheng and Y. Zhang, "Two-stage color ink painting style transfer via convolution neural network," in *2018 15th International Symposium on Pervasive Systems, Algorithms and Networks (I-SPAN)*, pp. 193–200, Yichang, China, 2018.
- [9] W. Jiang, Z. Wang, J. S. Jin, Y. Han, and M. Sun, "DCT-CNN-based classification method for the Gongbi and Xieyi techniques of Chinese ink-wash paintings," *Neurocomputing*, vol. 330, pp. 280–286, 2019.
- [10] J. Sheng and Y. Li, "Classification of traditional Chinese paintings using a modified embedding algorithm," *Journal of Electronic Imaging*, vol. 28, no. 2, p. 1, 2019.
- [11] J. Sheng, C. Song, J. Wang, and Y. Han, "Convolutional neural network style transfer towards Chinese paintings," *IEEE Access*, vol. 7, pp. 163719–163728, 2019.
- [12] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. S. Huang, and S. Yan, "Sparse representation for computer vision and pattern recognition," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1031–1044, 2010.
- [13] Y. Sun and R. Fisher, "Object-based visual attention for computer vision," *Artificial Intelligence*, vol. 146, no. 1, pp. 77–123, 2003.
- [14] R. Datta, *Semantics and Aesthetics Inference for Image Search: Statistical Learning Approaches*, [M.S. thesis], The Pennsylvania State University. ProQuest Dissertations Publishing, 2009.
- [15] V. N. Cudivada, "A geometry-based representation for efficient and effective retrieval of images by spatial similarity," *IEEE Transactions on Knowledge and Data Engineering*, vol. 10, no. 3, pp. 504–512, 1998.
- [16] A. Hanjalic, "Video and image retrieval beyond the cognitive level: the needs and possibilities," in *Storage and Retrieval for Media Databases 2001*, vol. 2001, pp. 130–140, San Jose, CA, United States, 2001.
- [17] L. Marchesotti, N. Murray, and F. Perronnin, "Discovering beautiful attributes for aesthetic image analysis," *International Journal of Computer Vision*, vol. 113, no. 3, pp. 246–266, 2015.
- [18] T. Matthews, M. S. Nixon, and M. Niranjan, "Enriching Texture Analysis with Semantic Data," in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1248–1255, Portland, OR, USA, 2013.
- [19] N. Mittal, D. Sharma, and M. L. Joshi, "Image Sentiment Analysis Using Deep Learning," in *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, pp. 684–687, Santiago, Chile, 2018.
- [20] P. Balouchian, M. Safaei, and H. Foroosh, "LUCFER: a large-scale context-sensitive image dataset for deep learning of visual emotions," in *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1645–1654, Waikoloa, HI, USA, 2019.
- [21] D. Kang, H. Shim, and K. Yoon, "A method for extracting emotion using colors comprise the painting image," *Multimedia Tools and Applications*, vol. 77, no. 4, pp. 4985–5002, 2018.
- [22] A. Yao, J. Shao, N. Ma, and Y. Chen, "Capturing Au-aware facial features and their latent relations for emotion recognition in the wild," in *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, pp. 451–458, New York, NY, USA, 2015.
- [23] D. Sen, S. Datta, and R. Balasubramanian, "Facial emotion classification using concatenated geometric and textural features," *Multimedia Tools and Applications*, vol. 78, no. 8, pp. 10287–10323, 2019.
- [24] Q. You, J. Luo, H. Jin, and J. Yang, "Robust image sentiment analysis using progressively trained and domain transferred deep networks," in *Twenty-ninth AAAI conference on artificial intelligence*, pp. 381–388, Austin, Texas, USA, 2015.
- [25] B. Jou and S. F. Chang, "Deep cross residual learning for multitask visual recognition," in *Proceedings of the 24th ACM international conference on Multimedia*, pp. 998–1007, New York, NY, USA, 2016.
- [26] V. Campos, B. Jou, and X. Giro-i-Nieto, "From pixels to sentiment: fine-tuning CNNs for visual sentiment prediction," *Image and Vision Computing*, vol. 65, pp. 15–22, 2017.
- [27] X. He and W. Zhang, "Emotion recognition by assisted learning with convolutional neural networks," *Neurocomputing*, vol. 291, pp. 187–194, 2018.
- [28] K. Song, T. Yao, Q. Ling, and T. Mei, "Boosting image sentiment analysis with visual attention," *Neurocomputing*, vol. 312, pp. 218–228, 2018.
- [29] Y. Fan, H. Yang, Z. Li, and S. Liu, "Predicting image emotion distribution by learning labels' correlation," *IEEE Access*, vol. 7, pp. 129997–130007, 2019.
- [30] T. Rao, X. Li, H. Zhang, and M. Xu, "Multi-level region-based convolutional neural network for image emotion classification," *Neurocomputing*, vol. 333, pp. 429–439, 2019.
- [31] W. Sun, H. Zhao, and Z. Jin, "A visual attention based ROI detection method for facial expression recognition," *Neurocomputing*, vol. 296, pp. 12–22, 2018.
- [32] X. Yang, Z. Wang, H. Deng et al., "Recognizing image semantic information through multi-feature fusion and SSAE-based deep network," *Journal of Medical Systems*, vol. 44, no. 2, 2020.
- [33] X. Liu, N. Li, and Y. Xia, "Affective image classification by jointly using interpretable art features and semantic annotations," *Journal of Visual Communication and Image Representation*, vol. 58, pp. 576–588, 2019.
- [34] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size," 2016, <https://arxiv.org/abs/1602.07360>.
- [35] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Image net classification with deep convolutional neural networks," in *Neural Information Processing Systems Conference*, pp. 1097–1105, Barcelona, Spain, 2012.
- [36] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, Las Vegas, Nevada, USA, 2016.

## Research Article

# Object Detection and Movement Tracking Using Tubelets and Faster RCNN Algorithm with Anchor Generation

**Prabu Mohandas** <sup>1</sup>, **Jerline Sheebha Anni** <sup>2</sup>, **Rajkumar Thanasekaran** <sup>1</sup>,  
**Khairunnisa Hasikin** <sup>3</sup> and **Muhammad Mokhzaini Azizan** <sup>4</sup>

<sup>1</sup>-Department of Computer Science and Engineering, National Institute of Technology Calicut, 673601, Kozhikode, India

<sup>2</sup>Department of Computer Science and Engineering, MEA Engineering College, Malappuram, -679325 Kerala, India

<sup>3</sup>Department of Biomedical Engineering, Faculty of Engineering, Universiti Malaya, 50603, Lembah Pantai, Kuala Lumpur, Malaysia

<sup>4</sup>Department of Electrical and Electronics Engineering, Faculty of Engineering and Built Environment, Universiti Sains Islam Malaysia, 71800 Nilai, Negeri Sembilan, Malaysia

Correspondence should be addressed to Prabu Mohandas; [prabu\\_pdas@yahoo.co.in](mailto:prabu_pdas@yahoo.co.in)

Received 9 April 2021; Accepted 20 July 2021; Published 10 August 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Prabu Mohandas et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Object detection in images and videos has become an important task in computer vision. It has been a challenging task due to misclassification and localization errors. The proposed approach explored the feasibility of automated detection and tracking of elephant intrusion along forest border areas. Due to an alarming increase in crop damages resulted from movements of elephant herds, combined with high risk of elephant extinction due to human activities, this paper looked into an efficient solution through elephant's tracking. The convolutional neural network with transfer learning is used as the model for object classification and feature extraction. A new tracking system using automated tubelet generation and anchor generation methods in combination with faster RCNN was developed and tested on 5,482 video sequences. Real-time video taken for analysis consisted of heavily occluded objects such as trees and animals. Tubelet generated from each video sequence with intersection over union (IoU) thresholds have been effective in tracking the elephant object movement in the forest areas. The proposed work has been compared with other state-of-the-art techniques, namely, faster RCNN, YOLO v3, and HyperNet. Experimental results on the real-time dataset show that the proposed work achieves an improved performance of 73.9% in detecting and tracking of objects, which outperformed the existing approaches.

## 1. Introduction

Elephants are pachyderms that live in the forest and move as groups in the search of food and water. Due to deforestation and climatic factors, elephant's movement in and around the forest areas has been increasing. These movements of elephants have led to problems such as elephants moving into human residing areas, elephants crossing roads nearby forest border areas, and crop raiding. As a result, the danger risk of human encountering the herds of elephants has become significantly dangerous, which may cause fatalities and destructions of human habitat. Therefore, there is an urgent need for

a technological approach in detecting and tracking the movement of the elephant herds. This paper looked at solving this challenge and proposed a methodology for elephant movement tracking and tried to find an optimal solution in detecting and tracking movements of the elephant. There have been several measures proposed such as electric fencing, elephant proof trench, acoustic detection, and image detection methods. However, these methods have certain disadvantages in tracking the elephant movement [1]. Through the video object detection methodology, the movement of elephant herds can be observed effectively. These herds, moving in between different groups, also become a factor in choosing

the object detection method [2]. On this approach, the analysis of the movements may also produce significant findings in terms of knowing their behavior and pattern of movement.

Video object detection is a technique involving object detection using video data compared to conventional object detection using static images [3]. Application areas of video object detection methods that have greater impact are autonomous driving and video surveillance. Video object detection approaches in the earlier stages have relied on manually analyzed features [4]. With the advancement in deep learning and convolutional neural networks, deep learning methods have been more effective than conventional approaches for various tasks in computer vision, speech processing, and multimodality signal processing. Specialized algorithms have been developed that can detect, locate, and recognize objects in images and videos, some of which include RCNN, RetinaNet, and YOLO.

In the proposed work, object localization and tracking are achieved using faster RCNN along with the tubelet generation. But using faster RCNN alone has the drawback of extracting similar features from the images when RPN is trained with minibatch size, and also, the network may need lot of time in the object detection process. However, object detection performance in faster RCNN requires further improved performance due to the problems in object detection such as occlusion and deformation. The proposed work overcomes the drawback of faster RCNN, through the framework faster RCNN with the tubelet generation method.

The elephant object in the images has been taken for analysis, which comprises of different patterns of object presence. Different patterns of objects in the images have been analyzed using the faster RCNN approach with anchor generation. Existing tubelet generation methods consider bounding box detection as object proposals and generate their own tubelet to track the objects. In the object tracking process, there has been a drifting problem leading to imprecise object location. To achieve precise localization, a tubelet generation method based on object detection has been proposed. Bounding boxes based on tubelet detection results in image object detection. During the frame detection in video, objects may be missed because of blur, artifacts, group object movement, etc. Hence, object detection and tracking had been made using tubelets for achieving the precise localization of the objects.

The primary contribution of the work includes the framework for (i) combining object detection in video frames and object tracking, (ii) object region proposals generated using faster RCNN for object classification while the tubelet generation method applied for object tracking, (iii) anchor generation method used with the faster RCNN process to predict the objects and its locations, (iv) real-time datasets collected from various forest areas had been used in the analysis, and (v) the elephant in the video frames considered for object movement and tracking.

## 2. Related Works

Significant research work has been made in the past for the object detection discussed in this section. Object detection and localization that had been made in recent years used

the still image. Video object detection methods have been effective in tracking the object localized. Hence, previous related works are analyzed for the object location and tracking.

Object detection in video has made great progress with neural networks and object detection algorithms. Still image object detections have been used effectively in refining object location in images, but it does not guarantee complete object instances in the image will be detected [5, 6]. The video object detection method has been applied in vast areas such as surveillance, transportation, and animal movement in forest areas [7, 8]. Selective search is an existing method for object detection, which generates box proposals for possible object locations by merging adjacent pixels in images [9, 10]. Object detection methods using only still images to detect objects lack accuracy because they cannot handle temporal and contextual information.

With the use of temporal in videos, object localization methods were proposed, which improves the object detection process. The object localization process is merely based on video frames similar to image object detection [11, 12]. Temporal consistency in the videos will have an impact in ensuring detection results for the video frames analyzed. Video analysis can vary significantly, such as human actions to object movement events [13]. In the existing approaches for video event, detection requires detecting and tracking objects initially, such as people, animals, and vehicles, then recognizing the actions of the objects.

Recognizing an object in the video is a developing area of research because of many fine-grained spatiotemporal variations [14, 15]. The objective of object classification is to find the object which appears in the video. In the proposed work, the problem of object localization and tracking is considered. The localized action detects changes in the spatiotemporal variations in a video.

*2.1. Object Detection.* Object detection is the process of detecting the bounding box which has the maximum score of detection for the given input image. Object detection in the video has been challenging due to varied image quality leading to unstable object classification in comparison to the object detection in static images. Using the tubelet generation method, the challenges in the video object detection method can be overcome by linking similar objects in the video to form tubelets [16]. The branch-bound method had been used for effective detection of bounding boxes [17]. Object detection methods such as still image detection, spatiotemporal, and contextual information in video were not explored completely [18, 19]. Hence, object detection methods combining still image and video will be effective.

Object detection performance has improved significantly with the deep neural networks. Neural network structures such as GoogLeNet, VGG, and ResNet were used to develop the learning capabilities on computer vision datasets for object detection, segmentation, and tracking [20, 21]. Neural network data such as images had been compressed during the transmission over the network and restored whenever required. It will help improve the detection accuracy [22]. Convolutional neural networks have shown improved performance in image analysis, especially in the areas of object

recognition and tracking [23, 24]. Bounding box proposals were generated from the image based on each location containing an object of interest [25]. Features extracted from each box proposal to classify it as one of the object classes. Feature extraction along with classification techniques will achieve low error rate in object detection [26]. Multiple networks had to be trained based on the different features extracted and bitrate compressions of the images taken for analysis [27]. Frameworks such as fast RCNN and faster RCNN formulate the object detection problem by training it on neural networks [28].

**2.2. Object Tracking Methodologies.** Object tracking is an important aspect in the process of locating the moving object in the video sequence. It is achieved by locating the target objects in consecutive frames and image pixels [29, 30]. To track the objects, object detection has to be made which has been attained using bounding box proposals in the proposed work. Machine learning approaches by extracting the features from video frames are used in tracking objects by locating the objects in the frames [31]. There have been different tracking algorithms such as Bayes, Euclidean distance, and intersection over union (IoU), for object tracking. The IoU algorithm has been efficient which involves finding the IoU between all combinations of objects of the current and previous frames. The IoU tracker can operate at thousands of frames per second, which outperforms other methods [32]. Accuracy and speed factor of the proposed work depend on the object tracker performance; hence, by using the object tracker like IoU, efficient results can be achieved.

Object detection in the video has been given increased attention due to the introduction of large datasets. Object detection in the video depends on the temporal information in the video. An efficient way to overcome this problem in the video is to analyze the temporal context of objects by linking the objects in the video to form tubelets. A tubelet used in object detection was defined as a series of bounding boxes associated with an object in image. In the proposed work, object proposals were formed by the region proposal network (RPN). The object proposals adopted in the video sequences will select the proposals between their neighboring frames with the scenario of the IoU overlap.

**2.3. Challenges in Existing Approaches.** Existing approaches have great success on detecting objects in static images, while detecting objects in videos remains a great challenge yet to be solved with great distinction. The challenges include factors such as drastic appearance, location change of the same object with the change in time, object occlusion, and motion blur [33]. In short, object detection approaches need to classify the object and also should be able to localize the objects in the video sequence. A previous method such as template-based action matching was used in object localization and classification [34]. Table 1 comprises of the notations used in the proposed approach.

### 3. Method for Object Detection and Tracking

To achieve object localization and classification, tubelet-based object detection with the faster RCNN was proposed.

TABLE 1: Notation with its description used in the proposed approach.

Symbol	Description
$S_t$	Tracking confidence of object
$S_d$	Detection score based on bounding box
$B$	Bounding box
$O$	Intersection over Union
$\beta$	Bounding boxes detected
$Tb_i, Tb_j, \text{ and } Tb_k$	Tubelet generated from video sequences
$a_1^i, a_2^i, a_3^i, \dots, a_w^i$	Visual features
$\Delta x_i^t, \Delta y_i^t, \Delta w_i^t, \Delta h_i^t$	Relative movement of objects
$W_w \text{ and } b_w$	Learning parameters
$b_i^j$	Bounding box locations
$m_1^i, m_2^i, m_3^i$	Object movement
$\{\tilde{M}\}, \{M\}$	Normalized movements of object
$I_1, I_2, I_3$	Video sequence
$O_t$	Object track
$D_n$	Object detections
$T$	Represents number of frames

The objective is to predict the high recall regions by detecting the objects in the image among the proposed regions. The detected region will be a background and objects from the given video set. Then, the model refines the localization and tracking of the object.

There is a need to detect and track objects such as elephants, due to its pattern of movement in different scenarios. Hence, the faster RCNN approach along with anchor generation has been proposed to detect the elephant's presence in different scenarios. Then, the elephant's movement had been tracked through the tubelet generation method.

The overall process of the proposed approach has been represented in Figure 1. Elephant detection and movement tracking has been made with the feature extraction and tubelet generation methods. Elephants must be initially detected to track the movement of the elephant in the video frames. Internal functions based on the feature extraction and tubelet generation methods were described in Figure 2.

The proposed approach for object detection and tracking is described in Figure 2. It includes an object classification process for classifying the objects detected in the video frames and a tubelet generation method for tracking the objects. For the given input video sequences, object proposals were generated. Based on the object proposal in the video frames, bounding boxes were determined in the object detection process. In the tracking process, objects have been classified and tubelets are generated.

**3.1. Object Localization and Classification Using Faster RCNN.** Object localization is to predict the object in the given video set. Similarities between the object locations were determined by the selective search approach, based on similarity criteria

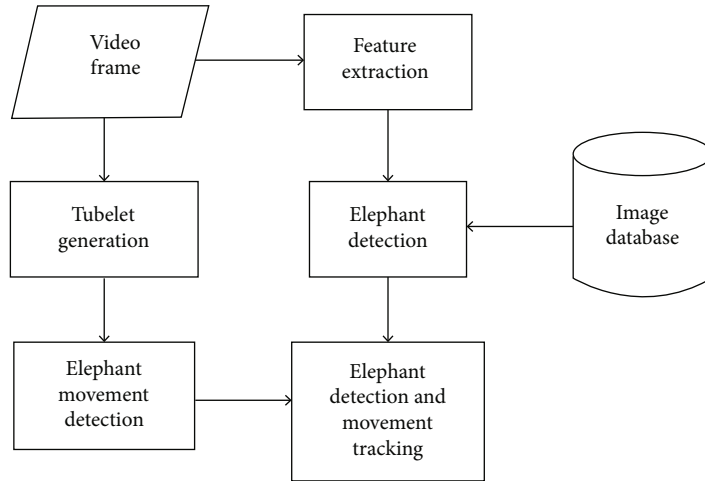


FIGURE 1: A proposed approach in a diagrammatic representation. After capturing the images of the elephant, feature extraction and tubelet generation methods were applied to track the elephant movement.

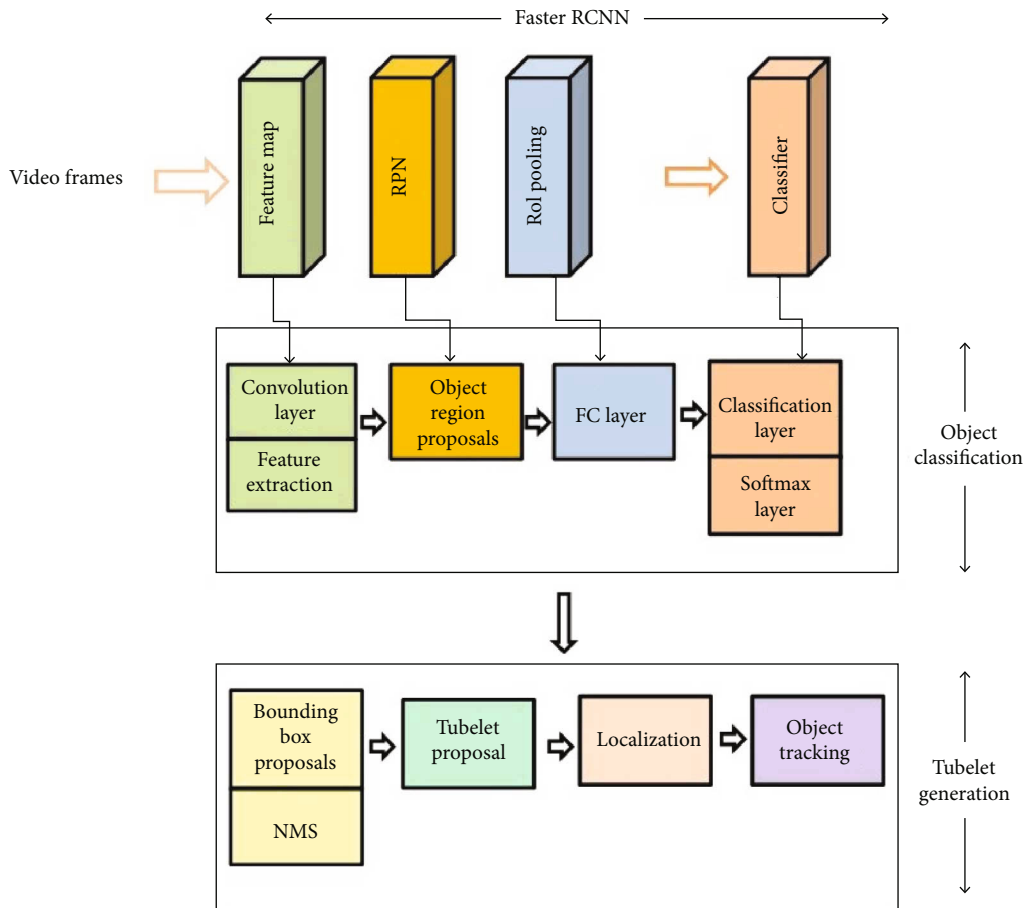


FIGURE 2: Block diagram for object detection and tracking. Object classification has been achieved through the layers such as the convolution layer, FC layer, and softmax layer. The tubelet generation method localizes the objects and tracks their movements.

for color, texture, and size. A selective search approach has limitations such as objects having different spatial locations within the image and varied aspect ratios. Hence, a large num-

ber of regions has to be selected which could result in a time-consuming process. To overcome these problems, the faster RCNN algorithm has been used to localize the objects and

improve the classification and detection process. The faster RCNN algorithm has significant improvement in the object detection process due to the region proposal network (RPN) which generates the object proposals [35]. RPN generates proposals for the objects in the images where the object exists.

Faster RCNN was used in the object detection task, which has two major functions such as generating region proposals and using these proposals in the network to detect the objects described in Figure 3. An input image is given to the convolutional layer which extracts the image pixels from the input image. Image pixels demonstrate the effectiveness of the image representation due to challenges such as corrupted input data [36]. Input consists of images arranged in the dimensions such as width and depth. Input holds the raw pixels of images with 3 color channels, consisting the feature map of an image [37]. Image prediction variations for the input images were based on the image pixel representations [38]. The convolution layer filters the image pixels, and the convolution operation is performed to attain a feature map. This is due to the fact that different objects were present in the images which had to be localized.

Elephant images captured in the forest areas of Theppakadu had been taken for analysis. The image containing an elephant object is passed to the convolution neural network in which faster RCNN generates object regions of interest (RoI). The next step has to pretrain CNN on image classification tasks, for defining the model. CNN in object classification takes input in the form of an image and provides the output as a category of the given images. CNN learns the feature along with the input data and uses two-dimensional convolution layer which is ideal for processing two-dimensional pictures. The region proposal network (RPN) was trained using the bounding box representation. RPN had to be fine-tuned for the regional proposal task which will be initialized by pretrain image classifiers. In the form of pixel coordinates, feature detection returns the region of interest. RoI will be a sequence of bounding boxes which is to be likely object positions. RPN is used in generating proposals, for the regions where the object is present. Feature maps are passed through a RPN for returning the object region proposals, which are classified further for object prediction and classification. Features are extracted from the images, are classified into different object classes, and return the bounding box. From the features extracted, the model was trained using the proposals generated by RPN. Then, faster RCNN had been used to initialize the RPN training in specific layers such as the convolution layer for object detection and classification.

**3.2. Object Classification and Localization.** The region proposal network in faster RCNN is given an input image and generates a set of object proposals for the corresponding feature map. The feature vector generated from the object proposals was fed into the output branches for object classification. In the last layer, object classification and localization were achieved. In the object detection task, each of the proposals will be of different shapes. Object proposals are detected to be different shapes based on the region of interest. Region of interest pooling converts the image proposals into

a fixed shape. Fixed sized feature maps are produced from nonuniform inputs by max pooling on the inputs.

For a given image, RoI pooling of each RoI depends on different parameters. It takes two inputs such as a feature map obtained from a convolutional network with convolution and max pooling layers and a matrix representing the regions of interest. The first column in the matrix denotes the image index, and the remaining column represents the coordinates of the object region. A fully connected (FC) layer has a softmax layer and a linear regression in which region proposals were passed for classifying and bounding box proposals for objects.

For the given input, based on the region of interest, a section of the input feature map is taken and scaled it to a fixed size. The scaling is done on the basis of dividing the region proposal into fixed size sections and finding the largest value of each section. Object localization in images was made using the similarity grouping of the nearby pixels. Similarities between the nearby pixels are acquired and merged with them. By repeated merging of the image pixels, the object location in an image was obtained.

The input image consists of an elephant as an object which was passed into a convolution layer to obtain the feature map as described in Figure 4. Then, the image filtered was passed through the RPN to obtain the localization of the objects which consists of elephant objects localized in the given input image. Object region proposals are of varied shapes and it has been normalized by the RoI pooling function. Hence, the objects will be of the same size located in an image. Then, the objects localized in had been classified. Based on the input image, the object has been localized and classified as elephant. The input image is of  $32 \times 32 \times 3$  describing the resolution and size of the image. Considering the given image in the JPG form with the dimension of  $320 \times 320$ , the representative array is of  $320 \times 320 \times 3$ . Numbers describe the pixel intensity which is of the value 0 to 255.

For the given input image, the feature map produced by applying the filter over the object locations of the image and an array of  $28 \times 28 \times 1$  was obtained as a feature map. 784 different object locations were obtained which can fit on to a  $32 \times 32$  input image. 784 object locations are mapped to a  $28 \times 28$  feature map. Filters perform a feature identifier function which includes things like edges and curves.

CNN has three main types of layers such as the convolutional layer, pooling layer, and fully connected layer. The convolution layer consists of filters which are small spatially that extends to the given input image. For the given input image, a 2-dimensional activation map that gives the responses of the filter for each position obtained, which means the network will learn the feature for the given input. There will be a number of filters in the convolution layer and an activation map will be produced by each of them. The output of the convolution layer will be combining the activation maps.

The pooling layer is to reduce the spatial dimension of the representation. It reduces the computation in the network and also controls overfitting. The pooling layer operates independently on the given input images and, by using the MAX operation, resizes the given input images. In the fully



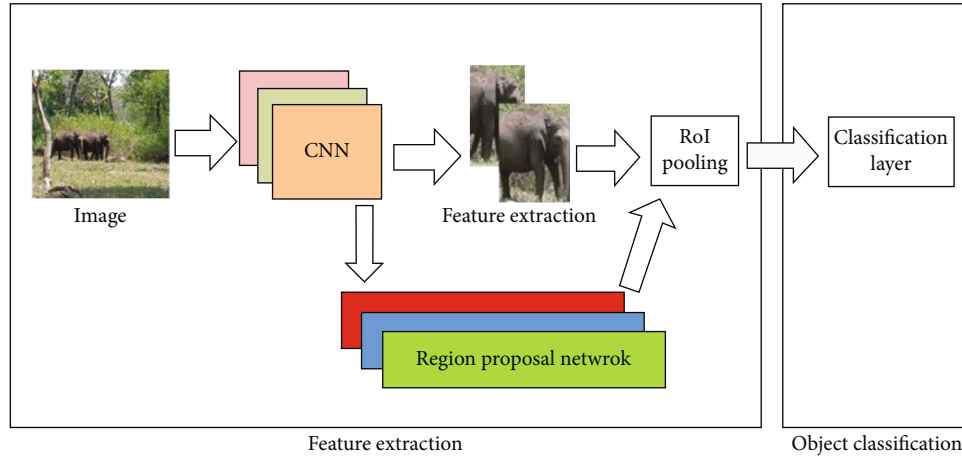


FIGURE 3: Faster RCNN for object detection. In the input images, CNN was used to extract the features, forming region of interest (RoI) pooling to classify the objects.

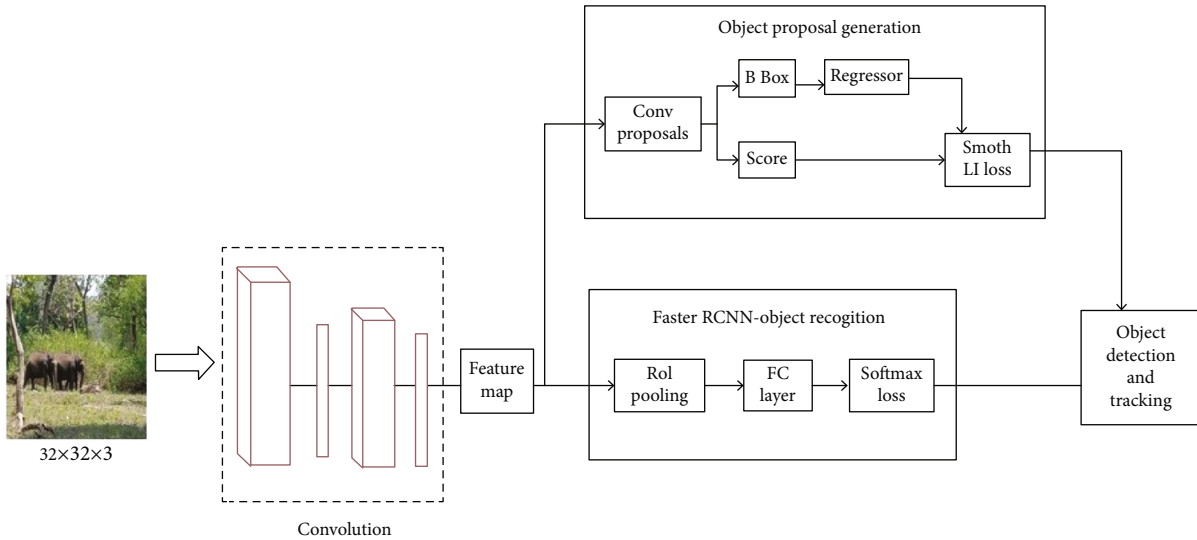


FIGURE 4: Proposed approach for object detection and recognition. Bounding box representation applied in object proposal generation for the elephant images acquired in the Theppakadu forest area. Objects were recognized using the faster RCNN approach.

connected layer, neurons have full connections to all activations in the previous layer.

The softmax loss layer determines the logistic loss of the softmax for the given input images. The function of the layer is to compute the logistic loss operation for different layers. The softmax layer has two inputs such as the predicted label and a fully connected layer. The smooth loss function does the classification performed with deep neural networks trained with the cross-entropy loss. Results suggest that cross-entropy is the learning objective for infinite data.

RPN generate the proposal for the objects. These object proposals are called bounding boxes. Bounding boxes were to determine the object locations in the domain by considering the ground truth box and IoU. RoI pooling in the object detection network resolves the problem of fixed size image requirement. The fully convoluted layer produces an output

of  $N$  dimensional object locations that has to be detected from the given input. Each of the objects represents the probability of an object location. Thus, the objects detected from the input image including trees and elephants are classified using the softmax layer.

**3.3. Object Generation Module.** In faster RCNN, RPN is used instead of the selective search module, which greatly improves the time and accuracy of object detection. Algorithms can also directly extract features to predict the object and its location. Methods such as OverFeat, YOLO, and SSD directly predict objects and return its location.

The proposed approach uses RPN to generate anchors used in direct classification and regress anchors. The number and shape of anchors will affect the object detection performance. Anchors are used in object detection for generating

object region proposals. More anchors are generated, leading to decreased detection accuracy, since most of the area surrounded by the anchor frame is the background. Hence, the anchor generation strategy can effectively reduce the number of anchors and will have great importance in optimizing the object detection performance.

In faster RCNN, the input image is passed through the convolution layer for anchor generation and feature extraction. Instead of RPN in the region, proposal generation and anchor generation have been proposed for determining the object presence. Object proposals generated from the anchor generation and features extracted from the feature map will be combined through RoI pooling to classify the objects. As described in Figure 5, initially, the anchor position prediction was made to generate a probability distribution in the feature map. It also indicates the possible location of the object in the image. In a similar manner, the object prediction involves prediction of the presence of objects in the input image. By combining the results of object and location prediction, the generated predicted objects in the images will be classified.

A probability distribution map generated by the anchor position prediction will be similar to the feature map of the input image. The proposed method is involved in determining the ground truth box prediction for the feature map in the training process and the threshold set for the remaining areas for determining the disregard area. The determination of the object prediction is to predict the length and width for the given anchor center point.

Anchor box characteristics are to capture the scale and aspect ratio of object classes required to be detected and chosen based on object sizes in the datasets. In the object detection, the anchor boxes were predefined and arranged across the image. Then, the network predicts the probability of object presence, background, and IoU. Predictions were used to determine each anchor box. In the proposed method, faster RCNN used to generate anchors with suitable size. The parameters  $(x, y)$  are used to describe an anchor, where  $(x, y)$  signifies the coordinates of the anchor. The shape information of the anchor had been integrated into feature map through which object detected can adapt to the anchor box parameters. As defined in Figure 5, a  $3 \times 3$  deformable convolution to the feature map, the offset of the convolution had been obtained through a  $1 \times 1$  convolution.

**3.4. Object Detection with Tubelet Generation.** Object detection with tubelet generation was made for the localization of the tubelets. Bounding box representation of the objects in the previous and current frame was compared to detect the presence of the same object in the two frames. IoU was used to predict the object location between different frames by using the threshold greater than 0.5, where the bounding box has to be the same. Lower detection scores may result during the bounding boxes tracking when the objects have larger overlap.

**3.4.1. Detection Scores.** Detection scores were calculated based on the bounding box coordinates and let the coordinates be  $(x_1, y_1)$  and  $(x_2, y_2)$ ,

$$S_t = \frac{1}{1 + e^{-O}}, \quad (1)$$

where  $S_t$  denotes the tracking confidence and with  $S_d$  detection score; object tracking has been achieved.  $O$  is the resultant of the average output for the given image.

Real-time elephant movement images as shown in Figure 6 are taken for a study from Hosur areas. Bounding representation of the object with its left, top, and bottom right coordinate representation had been described in Figure 6.

$$S'_t = \left\{ \begin{array}{l} (S_t + S_d) * \frac{1}{8} \\ (S_t + S_d) * \frac{2}{8} \\ (S_t + S_d) * \frac{3}{8} \end{array} \right\}, \quad (2)$$

where  $S_d \in [0, 0.25], [0.25, 0.5], [0.5, 0.75]$  and  $S_t$  is obtained by using the sigmoid function on tracking for ensuring the object presence based on the true objects detected as bounding boxes in expression (2). The values of the bounding boxes as stated in expression (2) are considered for analyzing the tracking of bounding boxes. In general, bounding boxes were considered true objects by the object tracker without analyzing the detection scores. The lower the detection score, the less likely to be the objects presence; hence, the tracking confidence and detection scores were calculated. There may be different bounding boxes which may have overlap within the frame. IoU was calculated, and if their value is higher than 0.5, then it is merged, since the detected boxes will be of similar objects.

In general, tubelets tend to overlap with each other due to multiple object detection as described in Figure 7 of an elephant image captured from Hosur forest areas. To overcome these types of problems, tubelets which are overlapping satisfying the above condition of value higher than 0.5 were merged. In the process of object tracking, object proposal suppression was performed to minimize the redundant tubelets. The objective of tubelet box distressing and max pooling process is to generate new tubelet boxes by replacing the existing tubelets on each frame randomly.

Figure 8 comprises of the object movement proposals for object classification in images with tubelets. Figure 8(a) is the input image comprised of an object classified, while Figure 8(b) defines the tracking of objects in an image. In the objects detected, tubelets had been replaced with those that have overlaps based on the threshold to perform the conventional NMS process. This process will bring back the positive boxes, if the tubelet detected has been with a lower detection score of positive boxes.

Object tracking and localization have been achieved using the bounding box sequence generated for the given video frames. A bounding box overlap for multiple objects had been determined using IoU overlap. Based on the IoU threshold, object movement has been tracked as described in Algorithm 1.

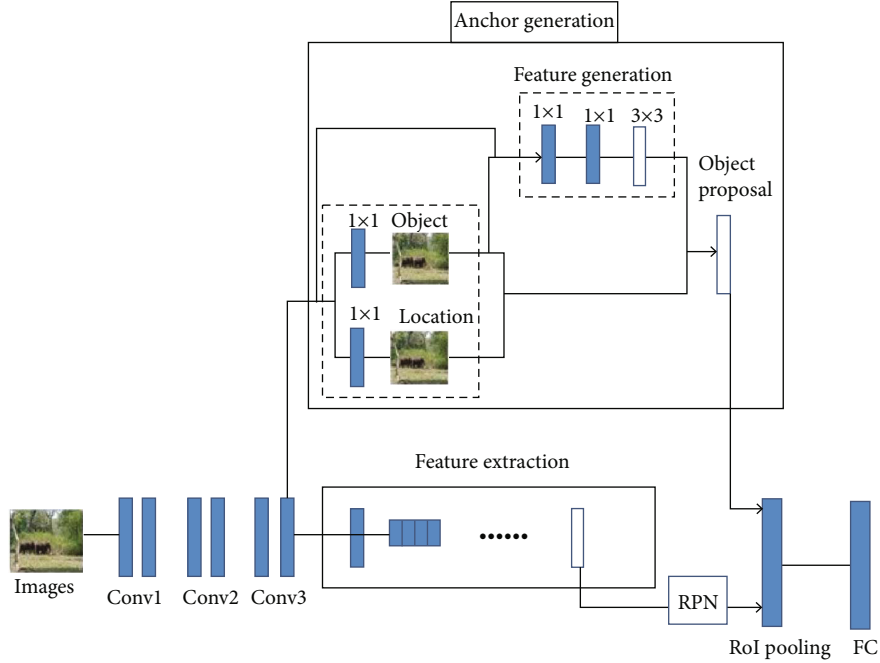


FIGURE 5: Object generation module using anchor generation to predict objects and location. Elephant images acquired in the forest area of Theppakadu were passed through convolution layers to classify the objects by anchor generation and feature extraction.



FIGURE 6: Bounding box representation for the input image. Bounding boxes of the object in the image with the coordinates were represented.



FIGURE 7: Intersection over union between objects. Bounding boxes were merged based on IoU values considering it as a similar object.

Tubelets were generated from the bounding box sequence for the given object instances as described in Figure 9.  $Tb_i$ ,  $Tb_j$ , and  $Tb_k$  were the different tubelets generated for the input video sequence.

**3.5. Object Tracking Based on Tubelet Generation.** There had been drifting problems in the object detection, due to bounding boxes overlapping in the detections. To overcome the drifting problems and precise object localization, the tubelet generation method was proposed. Object tracking was made based on the features of moving objects, by achieving prediction and tracking of moving objects. In the object tracking based on classification, a number of candidate regions will be extracted. These candidate regions will be sent to the network for classification and result in considerable time computation overhead. Object detections were represented in the form of bounding boxes. Tubelets are generated by applying the tracking algorithm to the static image bounding box proposals.

The proposed approach computes the object locations and classification scores for each frame. The classification score of the tubelet has been computed by combining the classification scores of the objects.

Tubelet proposals detect the objects in the images and can accurately track objects.  $a_1^i, a_2^i, a_3^i, \dots, a_w^i$  are visual features based on the box locations  $b_t^i$ . In order to track the movement of the object, a network is trained that effectively evaluates the spatial actions with respect to spatial features.

$$m_1^i, m_2^i, m_3^i \dots m_w^i = r(a_1^i, a_2^i, a_3^i \dots a_w^i). \quad (3)$$

Expression (3) was used to evaluate the movement of the



FIGURE 8: Detection of elephant object movement in the Theppakadu forest area with the bounding box proposal. (a) Object classification using tubelet generation. (b) Object tracking approach using tubelet for moving objects.

```

Step 1: input:  $B = (BB_1, BB_2, BB_3 \dots \dots \dots BB_n)$ ,  $S_d$ ,  $O$ ,  $\delta$ ;  $B$ —bounding boxes,  $S_d$ —detection scores of bounding boxes,  $O$ 
—intersection over union,  $\delta$ —NMS,  $BB_n$ —bounding box numbers.
Step 2: output:  $\beta$ —bounding boxes detected
Step 3:  $\beta \rightarrow \emptyset$ 
Step 4: while  $B \neq \emptyset$ 
Step 5: do
Step 6:  $\arg \max U(bb_j) \rightarrow bb_m$ 
Step 7:  $S_d(bb_m) \rightarrow S$ 
Step 8: if  $\text{IoU}(bb_m, bb_j) > \delta$  then
Step 9:  $S \rightarrow \max(S, S_d(bb_j))$ 
Step 10: end if
Step 11:  $\beta \rightarrow \beta \cup (bb_m, S)$ 
Step 12: end while
Step 13: return  $\beta$ 
    
```

ALGORITHM 1: Object tracking and localization using intersection over union (IoU).

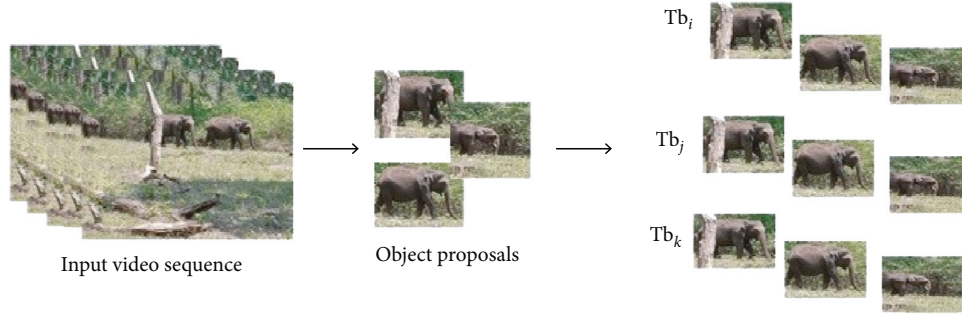


FIGURE 9: Tubelet generation from object proposals of the elephant movement images captured from Theppakadu forest area. Object proposals were generated by bounding boxes. Sequence of bounding boxes form the tubelets.

objects with the use of visual features. Relative movements are  $m_t^i(\Delta x_t^i, \Delta y_t^i, \Delta w_t^i, \Delta h_t^i)$  which has been calculated as follows:

$$\Delta x_t^i = \frac{x_t^i - x_1^i}{w_1^i}, \quad (4)$$

$$\Delta y_t^i = \frac{y_t^i - y_1^i}{h_1^i}, \quad (5)$$

$$\Delta w_t^i = \log \left( \frac{w_t^i}{w_1^i} \right), \quad (6)$$

$$\Delta h_t^i = \log \left( \frac{h_t^i}{h_1^i} \right). \quad (7)$$

By using (4)–(7), the relative movements of objects in the images can be inferred. The input taken as the visual features includes  $(a_1^i, a_2^i, a_3^i \dots a_w^i)^t$  and outputs the movement features of objects as  $4W$  expressed by the tubelet proposal as

$$(m_1^i, m_2^i, m_3^i \dots m_w^i)^t = W_w (a_1^i, a_2^i, a_3^i \dots a_w^i)^t + b_w, \quad (8)$$

where  $W_w$  and  $b_w$  are the learning parameters of the

```

Step 1: input: video sequence  $(I_1, I_2, I_3 \dots \dots \dots I_n)$ 
Step 2: data:  $O_t$ —object tracks,  $O_t = \{O_1, O_2, \dots, O_n\}$ ,  $D_n = \{D_1, D_2, \dots, D_n\}$  where
 $D_n$ —object detections,  $T$ —represents number of frames detected in the tracking and frames not detected in the tracking
Step 3: output: object track,  $O$ 
Step 4: Function tracking
Step 5: while  $\max A_{ij} > \alpha // A_{ij}$ —object tracks with intersection over union values
Step 6: do
Step 7: if  $B_{ij} \cup A_{ij} > \alpha // B_{ij}$ —number of frames in the video sequence without any object detections
Step 8:  $\arg \max A_{ij} \rightarrow (i, j)$ 
Step 9:  $D_j \rightarrow O_i$ 
Step 10:  $B_j \rightarrow True$ 
Step 11:  $O_t \rightarrow O_{t-1} \cup \{B_{ij} = True \forall D\}$ //object tracking for the detected objects
Step 12: return  $O_t$  for the given set of video sequences

```

ALGORITHM 2: Object tracking algorithm for the given video sequence.

TABLE 2: Dataset overview.

	Training	Validation	Testing
Video sequences	2758	684	2040
Number of frames	6329	4831	—
Positive snippets	1344	863	—

TABLE 3: Detection results for elephant object in video sequences.

Range	Detection results
(64, 128)	75.9%
(128, 256)	88.1%
(256, 512)	93.2%

TABLE 4: Object detection results for the given dataset.

Methods	Detection rate
TCNN	78.8%
Background subtraction	81.6%
Proposed approach	85.7%

TABLE 5: IoU overlap thresholds for different methods.

Method	0.25	0.50	0.75
Tubelet detection method	0.56	0.49	0.41
Static image video dataset	0.52	0.47	0.39
Proposed approach	0.71	0.63	0.54

layer. Tubelet boxes are generated by the regression layer that has similar movement patterns with the ground truth. The relative movement targets  $m_1^i = (x_i^t, y_i^t, w_i^t, h_i^t)$  can be defined on the basis of ground truth boxes at time  $b_t$ . Movement patterns of the objects with respect to the bounding boxes are represented through

$$L(\{\tilde{M}\}, \{M\}) = \frac{1}{N} \sum_{i=1}^N \sum_{t=1}^w \sum_{k \in (x,y,w,h)} d(\Delta k_t^i), \quad (9)$$

where  $\{\tilde{M}\}, \{M\}$  are the normalized movements for the object movement detection outputs. Object tracking based on tubelet generation primarily depends on the detection scores of a frame. Tracking of objects in images were based on the confidence of the object proposals. Anchors are called starting detections of tracking, based on the object proposals. There may be a drift in the objects detected during the object tracking moves away from the anchors. Hence, false tracking can be reduced by stopping the early object tracking while the confidence is below the threshold.

Algorithm 2 describes the process of object tracking in the video sequence. Object tracking represent the tubelet of the objects detected in each frame with the bounding box proposals. In the given video sequence, a set of tracking has been detected and based on the intersection over union measure, and tracking objects were identified. By detecting the objects in the frames, object tracking has been achieved.

## 4. Datasets

In the proposed approach, a new datasets consisting of real-time video recordings has been used for analysis. A dataset consisting of video sequences is captured during the field visits made at different seasons. During the field visits, elephant movements in different seasons and patterns were observed and recorded manually for future reference. Videos were captured in the format of  $1920 \times 1080$  pixels at 50 frames per second. The complete dataset comprises of 70 GB of video files, which approximately corresponds to 8.5 lakh frames. To avoid computational complexity in video processing, the video sequence taken for analysis is between 3 and 5 seconds, which includes major factors that may be available in the larger video sequences. For the evaluation of the proposed method, a subset of the video collection was taken for the study. Datasets taken were real time consisting of multiple objects. Images comprising multiple moving object categories were chosen, such as elephant, vehicles,

TABLE 6: Elephant object localization and tracking for the given dataset.

Method	Object localized
Tubelet detection method	66.3%
Tracking and detection based tubelet (proposed approach)	73.9%
Static image video dataset	71.8%

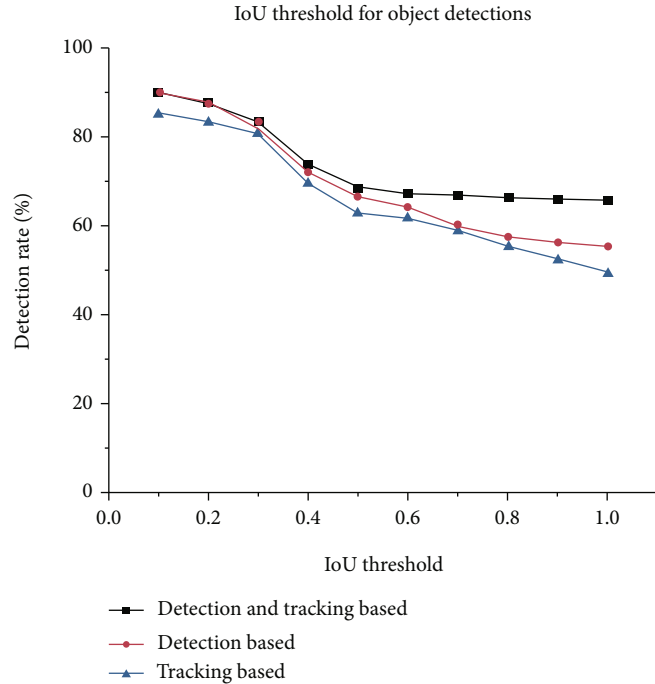


FIGURE 10: Object detection based on the IoU threshold. Performance comparisons of object detection and tracking methods based on the IoU were described in the figure.

and humans. In each input video, the numbers of frames are different.

Table 2 comprises of real-time datasets used in the process. The dataset is split into a training set and a validation set, containing 2758 video snippets and 684 video snippets, respectively. This mainly includes the category information of all objects which are identifiable and location information. In addition, the proposed object detection model performance was verified on the standard test set that contains 2000 video sequences and calculated the evaluation indicators in the test set.

**4.1. Evaluation Metrics.** Evaluation metrics were based on the average of precision on different aspects of object detection. The mean average precision was computed based on the score of a tubelet. The tubelet score was based on the average score of detection in the object detection process. Faster RCNN detectors were trained on different real-time datasets and network structures. Objects will be localized exactly based on the ground truth tubelet from a given class of images. An anchor has been considered a positive sample if it satisfies the constraints such as an anchor having the highest IoU overlap measure and an anchor having IoU greater

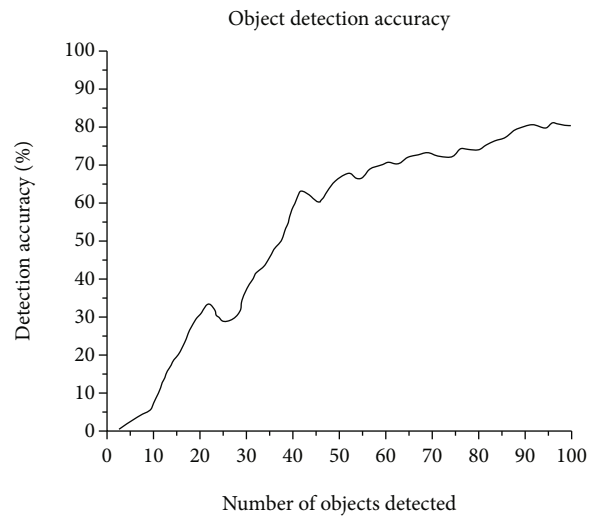


FIGURE 11: Object detection accuracy based on the number of objects. With the increase in the number of objects presence, the proposed method achieves a high detection rate.

TABLE 7: Performance comparison of the proposed approach with the existing approaches.

Methods	No. of frames	Specificity (%)	Precision (%)	Recall (%)	Accuracy (%)	TP rate (%)	FP rate (%)	Error detection rate (%)
Faster RCNN	572	0.7384	0.8513	0.8411	0.9235	0.9185	0.8331	0.9076
Kai et al. [25]	463	0.7572	0.8358	0.8824	0.9088	0.8938	0.8578	0.9012
Mihir et al. [9]	546	0.7264	0.8488	0.8517	0.9124	0.8754	0.8841	0.0876
Proposed approach	575	0.7841	0.9133	0.9357	0.9788	0.9341	0.9544	0.0212

TABLE 8: Object detection performance of the proposed approach.

Methods	Vehicles (%)	Monkey (%)	Deer (%)	Elephant (%)	Bear (%)	Humans (%)	Tiger (%)	Dog (%)
Faster RCNN	0.7813	0.8518	0.8551	0.8419	0.7915	0.8845	0.8688	0.8314
YOLO v3	0.8051	0.8493	0.8918	0.8588	0.8344	0.9145	0.8755	0.8489
HyperNet	0.7988	0.8369	0.8754	0.8635	0.7889	0.8801	0.8655	0.8581
Proposed approach	0.9311	0.8950	0.8845	0.9798	0.9318	0.9841	0.8845	0.8765

than 0.5 of the ground truth boxes. Anchors having IoU threshold less than 0.5 were considered a negative sample, and it have been ignored.

## 5. Results and Discussion

In the proposed approach, RPN anchors contained three scale (128, 256, and 512) models which were trained. The detection results of this proposed method reach 85.7%. The detection rate has been improved by exploring the model ensemble and computing the detection in different object ranges. Missing objects had been recalled by comparing it with the detection results which has been improved by the model ensemble as found in Table 3.

Based on the detection results and bounding boxes, tubelets are generated for tracking the objects. Object localization depends on the IoU threshold for the bounding box matching. IoU thresholds have multiple values for the bounding box matching such as 0.25, 0.5, and 0.75. The final tubelet generated will be based on an average of the tubelet mAP with different tubelet IoU thresholds. The tubelet detection was based on the actual tubelet localized to the given input video.

$$\text{IoU}(T, T_{at}) = \frac{T \cap T_{at}}{T \cup T_{at}}. \quad (10)$$

IoU thresholds were determined using expression (10) for object localization.

From Table 4, the object detection performance has been compared and the results show that the proposed approach has achieved higher performance. Detection comparisons have been made on the real-time dataset chosen for analysis. By using detection-based tubelet and tracking-based tubelet, object movement has been tracked in the videos.

*5.1. Threshold.* IoU is an evaluation metric used for describing the object detection model in the datasets. IoU evaluates

the bounding boxes predicted by the proposed model. IoU overlap thresholds for the different methods are presented in Table 5. These thresholds were determined based on the experimental results of object detection during overlapping. Based on the overlapping threshold, the object detection process for the given image varies. Threshold values such as 0.25, 0.50, and 0.75 were taken for considering the object detection variations on the images taken for analysis. With the increasing overlapping thresholds, the elephant object detection rate decreases. Increasing the overlapping thresholds of objects had resulted in varying detection results.

IoU overlap was set to the nonmaximum suppression (NMS) threshold to 0.3. Tubelets generated based on the bounding boxes in the video frames and their locations are predicted in the next frame using the optical flow value of bounding boxes. IoU overlaps were computed for each bounding box in the next frame. If the IoU overlap is of maximum value and above the threshold, then it belongs to the same object, else the bounding boxes represents a new object.

Table 6 compares the results of tracking and detection based on tubelet, to the other approaches for tracking and localization of objects in a video dataset. Based on the comparison, it is clear to define that the proposed approach outperformed the static image video dataset and tubelet detection method.

IoU thresholds for the object detection on the different scenarios were presented in Figure 10. Varying object detection rates were based on the bounding boxes for different kinds of tubelets. Detection- and tracking-based approaches were the proposed work, which had been effective in comparison with other approaches.

In Figure 11, the object detection performance based on the number of objects detected had been shown. When the number of objects to be detected is minimum, the detection rate has been higher. The minimum number of objects to be detected varies, since real-time scenario was given for the detection process. When the number of objects to be detected is less than 20, detection accuracy has been achieved above 80%.



FIGURE 12: Elephant movement tracking on the images and videos captured from the Hosur forest area and Theppakadu elephant camp by the proposed approach using tubelets. Elephants were detected in the first column of the images, and then, its movements were localized and tracked.

Tables 7 and 8 describe the performance of the object detection accuracy. The proposed approach performance has been described by comparing it with the existing approaches. Different objects detected using the real-time datasets have been tabulated below. From the results in Tables 7 and 8, it is conclusive that the proposed approach has been effective in object detection.

In Figure 12, an example of tubelet generation for elephant objects at different scenarios has been presented. Bounding boxes represent the object instance, and the similar object presence in the consecutive frames has been denoted by the same color. Hence, through the proposed approach, object instance detection and tracking have been achieved.



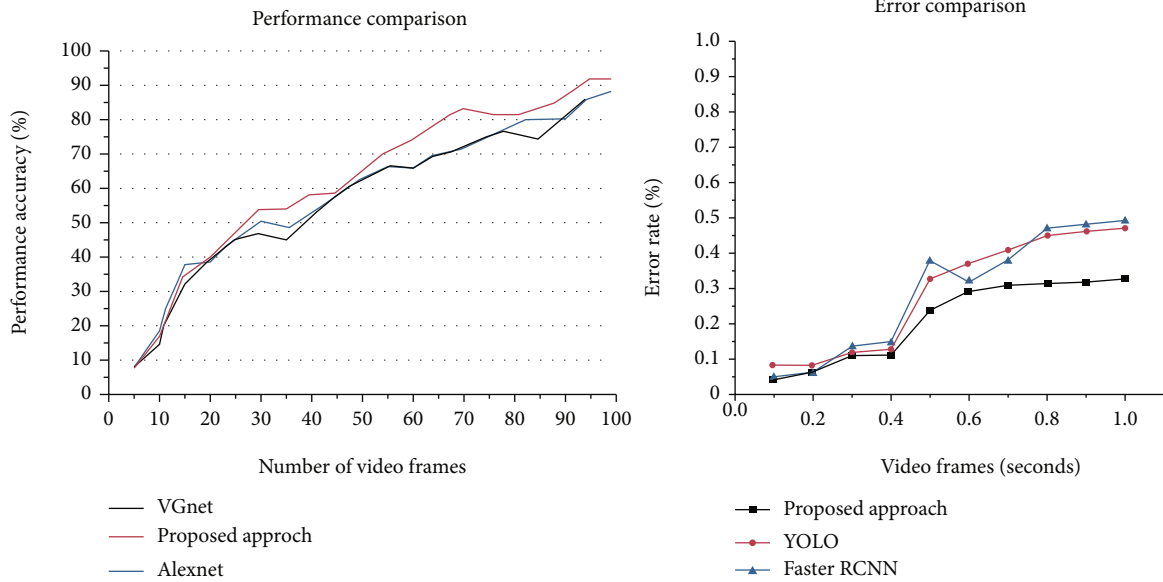


FIGURE 13: Performance and error rate comparison of the proposed approach. (a) Performance accuracy of the proposed approach compared with the existing approaches. (b) Error rate of the proposed approach compared with the existing approaches.

Figures 13(a) and 13(b) describe the performance and error rate comparisons of the proposed approach with the existing approaches. Based on the above comparison, it is clear that the proposed approach has been effective in object detection. The performance comparison of the proposed approach with other methods found data duplication due to the image pixel overlapping leading to misclassifications. Comparatively, other methods have been time consuming and require more memory space.

Faster RCNN has RPN to generate bounding boxes called region of interests (ROIs) which has high probability of containing objects. Hence, the number of bounding boxes generated indicates the presence of objects. The IoU threshold value was used in determining the object presence even though overlapping of bounding boxes exists which had been efficient in comparison to image pixel overlapping. In the proposed method, the object detection process had determined using a bounding box with a IoU threshold value. Then, the detected objects had been tracked through the tubelet generation method.

In the error comparison, localization errors had been found in the YOLO detections, and faster RCNN had a mean average precision (mAP) error variations in the object detections. The error rate in the object detections had been lower in the proposed approach while comparing to the existing approaches as described in Figure 13(b). Hence, the proposed approach will be an effective in the detection and tracking of elephants.

## 6. Conclusion

In this paper, the video object detection analysis is made through the machine learning approach. In the proposed approach, through faster RCNN and tubelet generation method, different objects in real time have been detected

and classified. The proposed approach has achieved 73.9% of detection and tracking of elephant objects which differs based on the image scale. Elephant objects have been tracked and classified using the IoU overlap of the anchor, where the different objects occlude. Using the detection results, elephant objects in the video had been localized and its movements had been tracked. The proposed approach has analyzed the elephant as an object for classification, and future work needs to be investigated on detection accuracy for multiple object detection. Furthermore, future study has to be made on object detection and tracking based on the object size variations.

## Data Availability

Data is available on authors request. Elephant images and videos were acquired from Hosur forest and Theppakadu elephant camp of Tamilnadu forest department.

## Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

## Acknowledgments

The work was supported and sponsored by the Department of Science and Technology, Science Engineering and Research Board (DST-SERB). The authors would like to express their gratitude to the forest officials and villagers from Hosur forest area and Theppakadu elephant camp for their help and support during the data collection procedure. The images of the elephant used in the paper are captured from the Hosur forest areas and from Theppakadu elephant camp. Images were captured and used with the consent from the authorities of Tamil Nadu forest department.

## References

- [1] A. Gayathri, A. Sulaiman, D. Kumar, S. Phalke, and A. Krishnan, "Status of elephant proof barrier mechanisms in Bannerghatta National Park," *Technical report*, A Rocha India, Bengaluru, 2016.
- [2] S. J. Sugumar and R. Jayaparvathy, "An improved real time image detection system for elephant intrusion along the forest border areas," *The Scientific World Journal*, vol. 2014, Article ID 393958, 10 pages, 2014.
- [3] H. Zhu, H. Wei, B. Li, X. Yuan, and N. Kehtarnavaz, "A review of video object detection: datasets, metrics and methods," *Applied Sciences*, vol. 10, no. 21, p. 7834, 2020.
- [4] H. T. Nguyen, M. Worring, and A. Dev, "Detection of moving objects in video using a robust motion similarity measure," *IEEE Transactions on Image Processing*, vol. 9, no. 1, pp. 137–141, 2000.
- [5] A. Arora, A. Grover, R. Chugh, and S. S. Reka, "Real time multi object detection for blind using single shot multibox detector," *Wireless Personal Communications*, vol. 107, pp. 1–11, 2019.
- [6] V. Rodrigo and R. D. S. Javier, "Object detection: current and future directions," *Frontiers in Robotics and AI*, vol. 2, pp. 1–7, 2015.
- [7] B. Benfold and I. Reid, "Stable multi-target tracking in real-time surveillance video," in *Computer Vision and Pattern Recognition*, pp. 3457–3464, IEEE, 2011.
- [8] C. Rd, J. Fernando, and G. Narciso, "An efficient multiple object detection and tracking framework for automatic counting and video surveillance applications," *IEEE Transactions on Consumer Electronics*, vol. 58, no. 3, pp. 857–862, 2012.
- [9] J. Mihir, V. G. Jan, J. Herve, B. Patrick, and G. M. S. Cees, "Tubelets: unsupervised action proposals from spatiotemporal super-voxels," *International Journal of Computer Vision*, vol. 124, pp. 287–311, 2017.
- [10] J. R. R. Uijlings, K. E. A. Sande, T. Gevers, and A. W. M. Smeulders, "Selective search for object recognition," *International Journal of Computer Vision*, vol. 104, no. 2, pp. 154–171, 2013.
- [11] O. B. Eshed and M. T. Mohan, "Multi-scale volumes for deep object detection and localization," *Pattern Recognition*, vol. 61, pp. 557–572, 2016.
- [12] M. Rochan, S. Rahman, N. D. Bruce, and Y. Wang, "Segmenting objects in weakly labeled videos," in *2014 Canadian Conference on Computer and Robot Vision*, pp. 119–126, 2014.
- [13] D. Jerline Sheebha Anni and S. Arun Kumar, "A wireless sensor network based on unmanned boundary sensing technique for minimizing human elephant conflicts," *Studies in Informatics and Control*, vol. 26, pp. 459–468, 2017.
- [14] P. Dollar, V. Rabaud, G. Cottrell, and S. Belongie, "Behavior recognition via sparse spatio-temporal features," in *2005 IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance (China, 2005)*, pp. 65–72, 2005.
- [15] J. Mao and L. Yu, "Convolutional neural network based bi-prediction utilizing spatial and temporal information in video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 7, pp. 1856–1870, 2020.
- [16] P. Tang, C. Wang, X. Wang, W. Liu, W. Zeng, and J. Wang, "Object detection in videos by high quality object linking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 5, pp. 1272–1278, 2020.
- [17] S. I. Mostafa, A. B. Amr, R. A. Mostafa, and F. E. Ibrahim, "Bounding box object localization based on image superpixelization," in *Proceedings of the International Neural Network Society Winter Conference (INNS-WC, 2012)*, pp. 108–119, 2012.
- [18] I. Everts, J. Gemert, and T. Gevers, "Evaluation of color spatio-temporal interest points for human action recognition," *IEEE Transactions on Image Processing*, vol. 23, no. 4, pp. 1569–1580, 2014.
- [19] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 9, pp. 1904–1916, 2015.
- [20] L. Li, O. Wanli, W. Xiaogang et al., "Deep learning for generic object detection: a survey," *International Journal of Computer Vision*, vol. 128, pp. 261–318, 2019.
- [21] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proceedings of International Conference Learn Represent (ICLR, 2014)*, pp. 1–14, 2014.
- [22] A. Namphol, S. H. Chin, and M. Arozullah, "Image compression with a hierarchical neural network," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 32, no. 1, pp. 326–338, 1996.
- [23] J. Y. Cheong and I. K. Park, "Deep CNN-based super-resolution using external and internal examples," *IEEE Signal Processing Letters*, vol. 24, no. 8, pp. 1252–1256, 2017.
- [24] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295–307, 2016.
- [25] K. Kai, L. Hongsheng, Y. Junjie et al., "T-CNN: tubelets with convolutional neural networks for object detection from videos," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, pp. 2896–2907, 2017.
- [26] W. Ouyang and X. Wang, "Joint deep learning for pedestrian detection," in *2013 IEEE International Conference on Computer Vision*, pp. 2056–2063, 2013.
- [27] C. Han, Y. Duan, X. Tao, M. Xu, and J. Lu, "Toward variable-rate generative compression by reducing the channel redundancy," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 7, pp. 1789–1802, 2020.
- [28] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 91–99, 2015.
- [29] S. V. Kothiya and K. B. Mistree, "A review on real-time object tracking in video sequences," in *2015 International Conference on Electrical, Electronics, Signals, Communication and Optimization (EESCO, 2015)*, pp. 1–4, 2015.
- [30] J. Li, B. Li, J. Xu, R. Xiong, and W. Gao, "Fully connected network-based intra prediction for image coding," *IEEE Transactions on Image Processing*, vol. 27, no. 7, pp. 3236–3247, 2018.
- [31] R. Christian, R. Yuan, and I. S. Nazrul, "Object detection and tracking in real time videos international," *Journal of Information Systems in the Service Sector*, vol. 11, no. 2, pp. 1–17, 2019.
- [32] B. Erik, E. Volker, and S. Tomas, "Training a convolutional neural network for multi-class object detection using solely virtual world data," in *13th IEEE International Conference on*

- Advanced Video and Signal Based Surveillance*, pp. 278–285, 2016.
- [33] W. Bin, T. Sheng, X. Jun-Bin, Y. Quang-Feng, and Z. Yong-Dong, “Detection and tracking based tubelet generation for video object detection,” *Journal of Visual Communication and Image Representation*, vol. 58, pp. 1–14, 2018.
- [34] L. Chengyou and T. Hua, “Human action recognition based on template matching,” in *Advanced in Control Engineering and Information Science*, pp. 2824–2830, Procedia Engineering, 2011.
- [35] Y. Xiao, X. Wang, P. Zhang, F. Meng, and F. Shao, “Object detection based on faster R-CNN algorithm with skip pooling and fusion of contextual information,” *Sensors*, vol. 20, no. 19, p. 5490, 2020.
- [36] K. Jia, X. Wang, and X. Tang, “Image transformation based on learning dictionaries across image spaces,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 2, pp. 367–380, 2013.
- [37] L. Yong-Hwan, K. Bonam, and K. Heung-Jun, “Efficient object identification and localization for image retrieval using query-by-region,” *Computers and Mathematics with Applications*, vol. 63, no. 2, pp. 511–517, 2012.
- [38] I. Schiopu and A. Munteanu, “Deep-learning-based lossless image coding,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 7, pp. 1829–1842, 2019.

## Research Article

# Simulation of Sports Venue Based on Ant Colony Algorithm and Artificial Intelligence

Rui Zhang <sup>1</sup>, Weibo Sun <sup>2</sup>, and Sang-Bing Tsai <sup>3</sup>

<sup>1</sup>Department of Physical Education and Research, Heilongjiang Bayi Agricultural University, Heilongjiang 163319, China

<sup>2</sup>Institute of Physical Education, Jiamusi University, Heilongjiang 154000, China

<sup>3</sup>Regional Green Economy Development Research Center, School of Business, WUYI University, China

Correspondence should be addressed to Weibo Sun; 80031124@qq.com and Sang-Bing Tsai; sangbing@hotmail.com

Received 8 June 2021; Revised 7 July 2021; Accepted 22 July 2021; Published 10 August 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Rui Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to improve the congestion of the evacuation plan and further improve the evacuation efficiency, this paper proposes the priority Pareto partial order relation and the vector pheromone routing method based on the priority Pareto partial order relation. Numerical experiments show that compared with the hierarchical multiobjective evacuation path optimization algorithm based on the hierarchical network, the fragmented multiobjective evacuation path optimization algorithm proposed in this paper effectively improves the evacuation efficiency of the evacuation plan and the convergence of the noninferior plan set. However, the congestion condition of the noninferior evacuation plan obtained by the fragmented multiobjective evacuation route optimization algorithm is worse than the congestion condition of the noninferior evacuation plan obtained by the hierarchical multiobjective evacuation route optimization algorithm. The multiple factors that affect the routing process considered in the probability transfer function used in the traditional ant colony algorithm routing method must be independent of each other. However, in actual route selection, multiple factors that affect route selection are not necessarily independent of each other. In order to fully consider the various factors that affect the routing, this paper adopts the vector pheromone routing method based on the traditional Pareto partial order relationship instead of the traditional ant colony algorithm. The model mainly improves the original pheromone distribution and volatilization coefficient of the ant colony, speeds up the convergence speed and accuracy of the algorithm, and obtains ideal candidate solutions. The method is applied to the location of sports facilities and has achieved good results. The experimental results show that the improved ant colony algorithm model designed in this paper is suitable for solving the problem of urban sports facilities location in large-scale space.

## 1. Introduction

When sudden disasters occur, the high-density people in the disaster area can be safely evacuated in the shortest possible time, and scientific and effective personnel evacuation strategies can be implemented to effectively reduce the serious consequences of the disaster [1]. The planning of the evacuation route is a key and important part of the evacuation. In order to be able to evacuate people in a specific area in a timely and effective manner, it is necessary to formulate an evacuation route plan [2]. If an efficient evacuation route plan can be implemented smoothly, it will be able to effectively reduce the evacuation time and shorten the length of the evacuation route. They reduce the degree of congestion and reduce casu-

alties and property losses [3]. Different evacuation scenes have different evacuation characteristics, and the exit position and internal structure of the evacuation scene have an important influence on the evacuation process [4]. Evacuation modeling based on the corresponding characteristics of the evacuation scene is beneficial to the improvement of the performance of the evacuation route plan. The open-air stadium has an approximate ring structure. The stand area is distributed around the inner ring, and multiple exits are distributed around the outer ring. There is a corresponding relationship between the stands and the VI. This structure determines that the evacuation of stadium personnel has the unique feature of evacuation from the inner ring stand area to the outer ring exit [5].

The current research on stadium personnel evacuation planning can be divided into path planning based on evacuation simulation and evacuation path planning from the perspective of optimization [6]. Most of the researches conducted from the perspective of optimization are single-objective evacuation route optimization or the conversion of multiobjective evacuation route optimization into single-objective evacuation route optimization. However, there are relatively few studies on multiobjective evacuation route optimization based on multiobjective optimization theory. Multiobjective evacuation path optimization often uses general multiobjective optimization algorithms, and there are fewer dedicated multiobjective optimization algorithms based on the evacuation characteristics of stadium personnel [7]. The general multiobjective optimization algorithm uses the pseudorandom ratio process to search the evacuation route, lacking the guidance of domain knowledge, and it is easy to fall into a blind search. A dedicated multiobjective optimization algorithm based on the evacuation characteristics of stadium personnel, with the help of domain knowledge, enhances the purpose of the search, reduces the scope of the search space, and easily obtains an evacuation route plan with higher evacuation efficiency and better evacuation performance [8].

In order to improve the evacuation efficiency of the evacuation plan and improve the evacuation performance, this paper presents the corresponding topological structure characteristics of the fragmentation for the stadium stand area and exit, abstracts the entire stadium into a fragmented network, establishes a fragmented network personnel evacuation model based on the characteristics of the stadium, and proposes a fragmented multiobjective evacuation path optimization algorithm based on this model. Using this structural feature, the stadium is abstracted into a fragmented network to guide the evacuation process of evacuated persons. In this paper, the problem of slow convergence of a traditional ant colony algorithm is improved. With reference to the idea of an artificial potential field method, a gravitational probability function with adjustable weights is constructed and added as a heuristic factor, so that the new algorithm can still converge at a faster speed. Under the guidance of the network, the evacuated persons can only evacuate to the exits located in their own shards, which prevents a long cross-shard path during the evacuation process, which is beneficial to shorten the length of the evacuation path and improve the evacuation efficiency.

## 2. Related Work

Ma and Tsai [9] put forward the problem of how to coordinate the contradiction of multiple goals, which is the embryonic form of the problem of a multiple 1:1 standard optimization. French economist Alatas (see [10]), from the perspective of political economy, summarized many incomparable goals into multiobjective optimization problems and gave the definition of multiobjective optimization problems for the first time. Gao et al. [11] gave the point of view of a game theory; it puts forward a multiobjective decision-making problem with multiple decisions that contradict each

other. Through the analysis of production and distribution activities, the multiobjective optimization problem is proposed, and the concept of the Pareto optimal solution is proposed for the first time. Tsai and Ma [12] gave the concept of the Pareto optimal solution to solve the vector extremum problem from the perspective of mathematical programming and analyzed the sufficient and necessary conditions for understanding. Since then, the problem of multiobjective planning has gradually attracted people's attention. After the 1990s, the multiobjective optimization problem formally became a branch of mathematics, attracting more and more scholars and experts in different fields to conduct systematic research, and it became a research hotspot. At present, the optimization of multistandard standards has formed a relatively complete theory and has obtained many important research results. Its theories and results have been widely used in various fields, showing its strong vitality more and more.

In terms of evacuation of people in buildings, a lot of research work tries to capture the evacuation behavior of people through the computer simulation of evacuation models and then uses experimental data to refine the proposed evacuation model. Researchers at home and abroad have proposed a number of evacuation models [13], such as Simulex, EXODUS, BGRAF, and EXITT, and formed pedestrian evacuation software: ACNET, SIMULEX, SGEM, HAZARD, EGRESSPRO, EXIT, STEPS, etc. In recent years, Ai et al. [14] have analyzed and summarized these models, including seven models and implementation methods: cellular automata models, lattice gas models, social dynamics models, fluid dynamic models, agent models, and game theory models and animals based on experiments. In analyzing the process of pedestrian evacuation, the cellular automaton model describes the environmental impact (such as friction and congestion) and pedestrian capabilities (such as path selection ability, coordination ability, action ability, and emergency experience). Objectively, they reflect the environment and pedestrian interaction process, as well as analyze the influence of obstacles, and realize the description and analysis of competitive exit behavior and two-way movement caused by human interference in emergency situations. The lattice gas model is a discrete-time system model. In each time unit, all pedestrians in the model are distinguished by random sequence rules. Kendall et al. [15] use a random selection function to complete the selection process of a random moving queue. Statistics and probability are commonly used to describe the characteristics of the flow of people. The lattice gas model can reflect the nature of the dynamic evacuation process and analyze the time and space distribution of the evacuation time of the flow of people. The mobile lattice gas model proposed by Fister et al. [16] reduces the computational intensity and supports the evaluation of the average evacuation time. However, since the collection of test data is more difficult, the pedestrian characteristic analysis in this model lacks an effective verification method.

In terms of vehicle evacuation on the road network, the evacuation can be divided into static network-based vehicle evacuation and dynamic network-based vehicle evacuation according to the nature of the road network [17]. The

problem of vehicle evacuation in static networks can be attributed to the shortest evacuation path, fastest evacuation flow, and maximum evacuation flow [18], for example, based on the traffic capacity of the road network, the evacuation path is studied based on the maximum flow method; Huang [19] studies the influence of pedestrians, vehicles, intersection travel time, and path selection changes on the evacuation time of the road network; based on the shortest path algorithm in graph theory, some scholars have proposed an optimal independent evacuation path method based on the  $K$  shortest path. The basis for determining the optimal evacuation path includes the maximum capacity and travel time on the road section; some scholars use the minimum cost flow problem for evacuation traffic allocation, the shortest evacuation plan method is proposed, and the shortest evacuation plan is obtained; there is also a dual-objective comprehensive optimization model for emergency evacuation, which strives to achieve the maximum passing capacity, realize the maximization of all the beneficial results during the evacuation period, and realize the optimization of the public evacuation time in stages; some scholars use the shortest route and different route algorithms to correct the basic information of the road network in complex situations such as public emergencies. The improvement of dissimilarity calculation makes the optimal path and dissimilarity path more in line with that of the traveler's [20]. Vehicle evacuation based on a static network can complete the optimal allocation of evacuation vehicles from a macro perspective, but the dynamic and unpredictable characteristics of the evacuation process make it impossible to implement this evacuation plan accurately, and the effect of the evacuation model is not ideal [21, 22].

### 3. Construction of Sports Venue Model Based on Ant Colony Algorithm and Artificial Intelligence

*3.1. Spatial Distribution of Sports Venues.* In single-objective optimization problems, the optimal solution is usually uniquely determined, while for multiobjective optimization problems, because the various objective functions conflict and cannot be compromised, the optimal solution set of multiobjective optimization problems is usually a set, which is the essential difference between a multiobjective optimization problem and a single-objective optimization problem. Figure 1 shows the spatial distribution of sports venues (see Figure 1).

In the multiobjective optimization problem, the Pareto dominate (dominate) is a very important concept; it is a partial order relationship of the solution, defined as follows: for any two decision vectors  $a$  and  $b$  in the decision space  $X$ ,  $a$  governs  $b$  if and only if the target value of vector  $a$  is not greater than the target value of vector  $b$ . The corresponding target value of the vector is strictly smaller than the target value corresponding to vector  $b$ .

$$x(t+1) = (1-t) \times x(t) + \Delta t. \quad (1)$$

The so-called optimal solution of the multiobjective optimization problem refers to the Pareto optimal solution, so all the Pareto optimal solutions constitute a set of reasonable solutions for the multiobjective optimization problem.

$$y(x) = \begin{cases} x(t)^\alpha * \eta(t)^\beta, & x < t, \\ 0, & x > t. \end{cases} \quad (2)$$

In most cases, there is no definite optimal solution in multiobjective optimization problems, only Pareto optimal solutions exist, and multiobjective optimization problems usually have multiple Pareto optimal solutions. The Pareto optimal solution of the multiobjective optimization problem can be regarded as an acceptable noninferior solution.

$$D(i, j) = \sqrt{(x(i) - X)^2 + (y(i) - Y)^2}. \quad (3)$$

*Pareto optimal solution:* for any vector  $a$  in the decision space  $X$ ,  $a$  is Pareto optimal, if and only if there is a decision vector of the Pareto dominance vector  $a$  in the decision space  $X$ .

$$u(i, j) = \frac{\sum |\Delta D(m, n)|}{\sum |\Delta D(m, n)| - D(i+1, j)}. \quad (4)$$

*Pareto optimal solution set:* it is composed of all Pareto optimal solutions in decision space  $x$  and defines the target vector set corresponding to this set as a Pareto Optimal Front.

$$v(i, j) = \frac{x(i+1, j)}{\sum x(m, n)} + \frac{k * D(i+1, j)}{1+k}. \quad (5)$$

Based on the above description of the characteristics of deep learning theory, a deep learning optimization ant colony algorithm suitable for project technical and tactical decision-making is established to solve the actual technical and tactical decision-making optimization problem.

$$w(i, j) = \begin{bmatrix} x_{11} & 0 & 0 \\ 0 & \cdots & 0 \\ 0 & 0 & x_{(i,j)} \end{bmatrix}. \quad (6)$$

The optimization process is as follows: (1) Set the number of iterations and the initialization of pheromone data. (2) Record the information of each confrontation process, including time and combination of skills and tactics to construct a taboo table for both sides' technical and tactical decision-making, use deep learning to conduct self-organized deep learning on the taboo table data, and use the output result decision-making combination taboo table to judge optimal technical and tactical decision-making route. (3) Construct the optimal solution: complete the technical and tactical changes and construct the technical and tactical decision-making route. The task of each ant is to find the optimal technical and tactical decision-making route. (4)

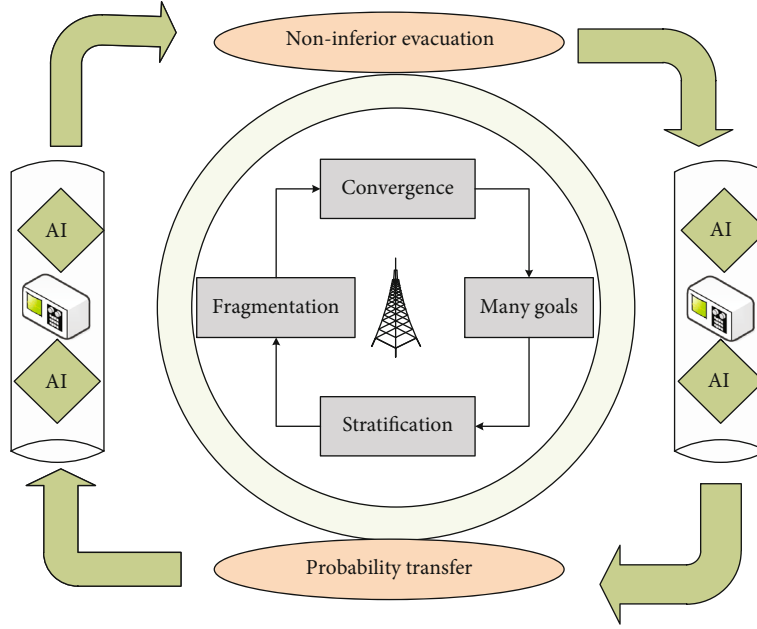


FIGURE 1: The spatial distribution of sports venues.

Update the data in the taboo table to determine whether the optimal victory has been achieved.

**3.2. Information Iteration of Ant Colony Algorithm.** The traditional basic idea of solving multiobjective problems is to convert multiobjective optimization problems into single-standard optimization problems, that is, to aggregate multiple subobjectives into a single-objective function with positive coefficients (weights) in the form of weighted summation and then use single-objective optimization. In the basic ant colony algorithm, the ants select the next node according to the transition probability, and this depends on the heuristic information of the candidate node and the information decay concentration, but the heuristic information has locality and cannot connect the current position with the evacuation exit. The pheromone concentration of the basic ant colony algorithm may cause the pheromone concentration to be too high as the ants gather on a certain path. The algorithm is solved, and the coefficients are determined by the algorithm designer or adaptively adjusted by the optimization method.

$$\begin{cases} \sum x(i, j) = 1, & i \in N, \\ \sum x(j, j) = p, & j \in N, \end{cases} \quad (7)$$

$$\frac{\partial u(x)}{\partial x} * \frac{\partial v(x)}{\partial x} = \frac{\partial z(x)}{\partial x}.$$

The solution method based on the Pareto dominance relationship can be divided into two stages according to the development process. The first-generation evolutionary multiobjective optimization algorithm combines Pareto ranking, fitness sharing, and niche strategies. The corresponding

algorithm has a multiobjective genetic algorithm (MOGA), which is not inferior.

$$[f1, f2, \dots, f(n)] * \begin{bmatrix} x1 \\ x2 \\ \dots \\ x(n) \end{bmatrix} = \sum \sum \min(\text{dis}(u(x)) * \text{road}(p)). \quad (8)$$

This kind of agent makes real-time decision-making at runtime, usually based on limited information and simple situation-action rules, which can establish a direct mapping between sensory information and action sets and has strong adaptability. Compared with the deliberate structure, the reaction structure can better meet the real-time requirements (see Figure 2).

Figure 2 shows the iterative process of ant colony algorithm information. We use a global high-level decision-making module to determine the team's current tactics, formation, and agent-to-role mapping and notify each player agent of the results of high-level decision-making, and then, the player agent will react according to its own knowledge and instructions from the high-level decision-making module. Agents on the field are divided into two states: strategic running state and active state. Agents in the strategic running state perform strategic moves according to the current team formation and their own roles; agents in the active state will perform reactive actions based on their surrounding environment. The sending module in the vision system puts the information and uses the TCP protocol to send it to the data forwarding server, and the data forwarding server sends the data to the decision-making system. The advantage of 4 partitions is that there are not many areas, the coding work is not

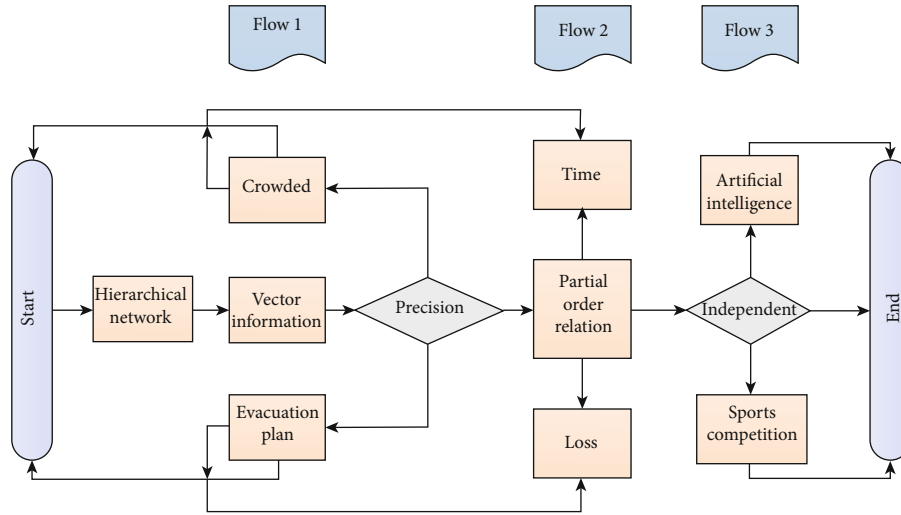


FIGURE 2: The iterative process of ant colony algorithm information.

heavy, and the jumping of the area is refined, which effectively prevents the sudden jump of the target. Zoning also reflects the obvious division of offensive and defensive areas, basically accommodating all specific situations. According to the division of these areas, and then considering the current distribution of robots, the ball-holding side, and the scores of both sides, the decision-making system can formulate the overall strategy of the team.

**3.3. Artificial Intelligence Level Optimization.** The evolutionary multiobjective optimization algorithm is an optimization algorithm characterized by the heuristic search of the population and the partial order relationship of the solution (such as the Pareto dominance relationship). Its appearance brings a new way to solve the multiobjective optimization problem [23–25]. The algorithm can not only solve the single-objective optimization problem well but can also obtain a set of compromise feasible solutions through one operation by using the Pareto dominance relationship. The decision maker can choose a suitable solution among the feasible solutions according to the corresponding domain knowledge and preference requirements. Since the perception range of the crowd is limited, new paths may be blocked by obstacles again. The walking stage of the entire crowd is a process of constantly updating the electronic map and planning new paths. The method is very suitable for solving multiobjective optimization problems. The essence of deep learning optimization is to learn more useful features in the data by constructing a machine learning model with a high-level hidden layer and massive training data and ultimately improve the accuracy of classification and prediction, including the depth of multiple hidden layers [26–28]. Figure 3 shows the comparison of histograms of site space units under different algorithms. In deep learning, it is different from traditional artificial neural network learning, which is reflected in the deep model structure and feature learning: (1) the depth of the model structure increases from the shallow three-layer hidden layer to 5, 6, and even 10 layers; (2) clear feature learning, through layer-by-layer feature training, the

learning samples are transformed from original features into new feature learning, making classification and prediction easier (see Figure 3).

With a hybrid structure, the entire decision-making process is divided into two layers. Figure 4 shows the framework of the sports venue model based on ant colony algorithm and artificial intelligence. The first layer is the team coordination layer based on reasoning. The decision-making information comes from the visual information sent by the vision server and the referee information sent by the referee box. It formulates the overall strategy of the team and determines the current team's formation and the role of each player. The role of the player is to coordinate the actions between the players; the second layer is the player action layer based on the reactive structure, according to the team's decision-making layer's formation and assigned roles; we select the player's actions and actions to be used in the next cycle (see Figure 4).

In the early stage of path planning, the pheromone concentration is insufficient. At this time, the potential field force is mainly used to inspire information. The role of the potential field force ensures the accuracy of the path planning direction, reduces the randomness of the search, and produces poor quality solutions. With the deepening of the search, the role of pheromone concentration and distance enlightenment is exerted, and the effect of the potential field force is gradually reduced to prevent the path from being excessively concentrated in the direction of the potential field gradient. The improved distance heuristic information function enables the search to advance to the target point with a high probability, which improves the efficiency.

## 4. Application and Analysis of Sports Venue Model Based on Ant Colony Algorithm and Artificial Intelligence

**4.1. Ant Colony Algorithm Data Collection.** The study area is dominated by the entire area, covering an area of 304.9 km<sup>2</sup>. The entire area is divided into 512 × 512 spatial units. The



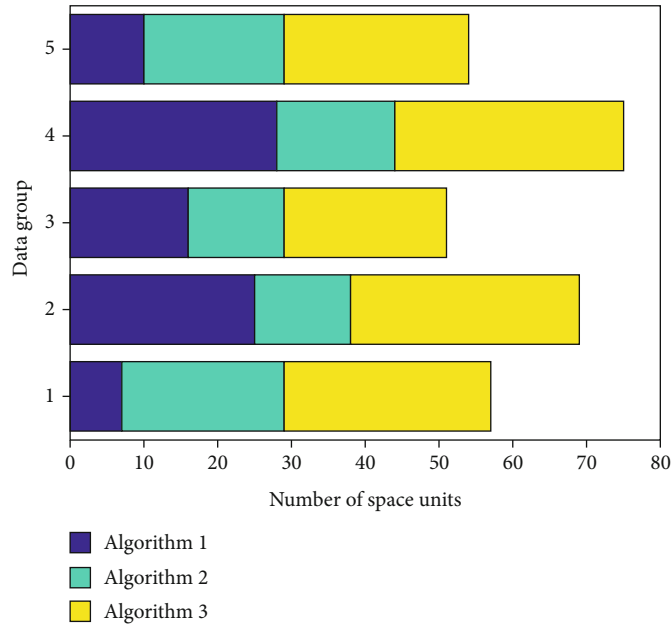


FIGURE 3: Comparison of site space unit histograms under different algorithms.

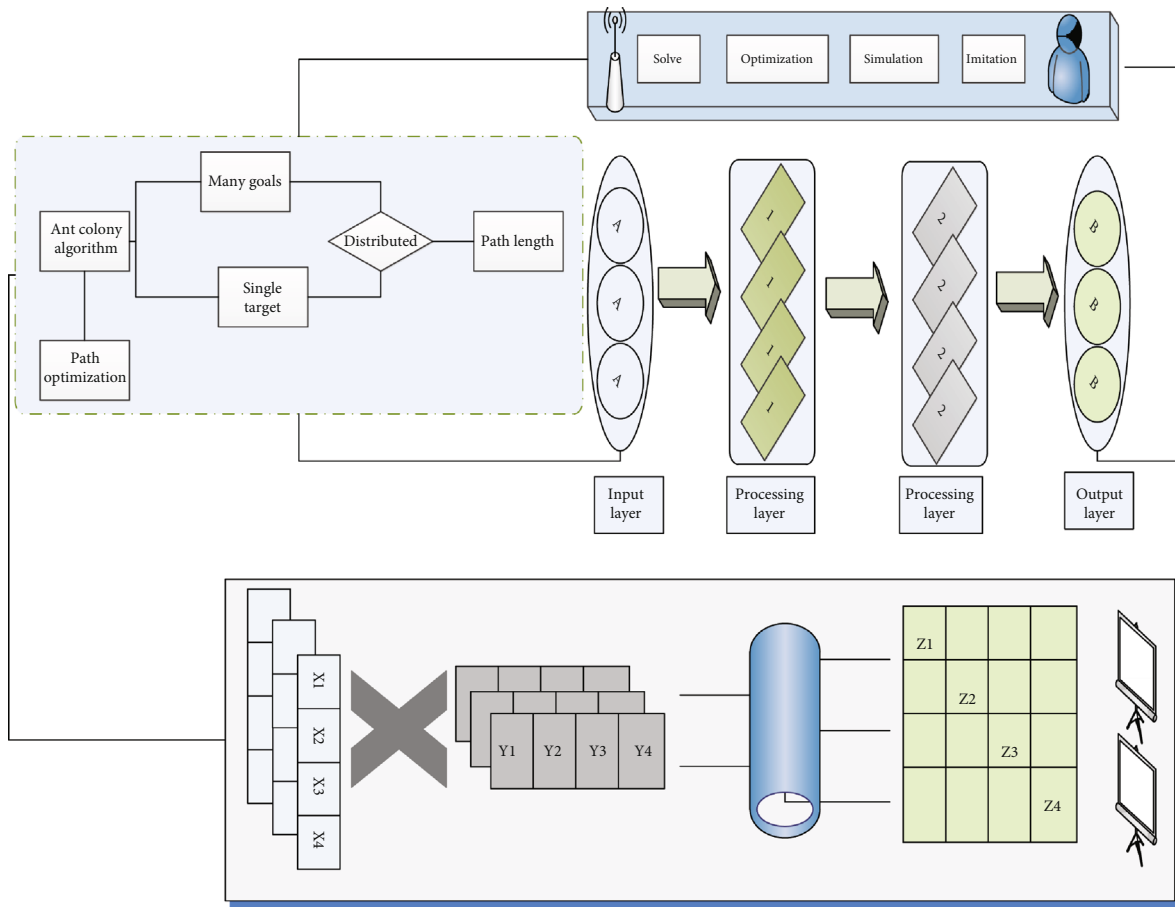


FIGURE 4: Sports venue model framework based on ant colony algorithm and artificial intelligence.

area of the unit grid is  $12 \times 12 m^2$ . In ant colony algorithm path planning, the number of ants  $m$  affects the algorithm's global search capability and convergence speed. Increasing

$m$  can increase the randomness of the search, but the convergence speed will slow down. If  $m$  is too small, the search will stop prematurely. In a grid environment of  $20 \times 20$ ,  $m = 20$  is

used in this paper. The pheromone factor  $Q$  is related to the positive feedback effect of the algorithm search. Increasing  $Q$  can increase the convergence speed, but too large  $Q$  can easily fall into the local optimal solution. Pheromone concentration volatilization coefficient  $\rho$ , pheromone concentration factor  $\alpha$ , and heuristic information factor  $\beta$  reflect the strength and convergence speed of the randomness of the algorithm search process. The best value range of the three parameters is  $0.1 \leq \rho \leq 0.99$ ,  $0 \leq \alpha \leq 5$ , and  $0 \leq \beta \leq 8$ . In this paper,  $\alpha = 1$ ,  $\beta = 7$ , and  $\rho = 0.3$  (see Figure 5).

The selected spatial data information mainly contains the vector data information of 12 streets. At the same time, ARC-GIS software is used to edit the geographic data information. Through editing,  $512 \times 512$  spatial raster data information can be obtained, which is further processed into data information in text format. Figure 5 shows the site selection curve for different data points. When the grid map is initialized, each node is given equal pheromones, which can expand the initial search range of the electronic ant as much as possible. Then, use the site selection model of this article for simulation calculation and visually display the calculated data results. Java language is used to implement the sports facility location model in this article. Before the model runs, the parameters in the system need to be set. The current setting of the ant colony algorithm initialization parameters is determined by empirical methods or a large amount of experimental data. According to the empirical value method, we set the pheromone intensity in this article as  $Q = 10000$ , the pheromone volatilization coefficient is  $\rho = 0.01$ , the number of ants in the ant colony is initialized to  $\text{ant\_nums} = 30$ , pheromone heuristic factor is  $\alpha = 1$ , and the pheromone expectation heuristic factor is  $\beta = 0.3$ . In this paper, the maximum number of repetitions of the ant colony algorithm  $M$  and the number of iterations  $P$  are used as the conditions for the termination of the algorithm, which can reduce the execution time of the algorithm. Through continuous experiments and tests, we can obtain sports facilities with different numbers of locations.

**4.2. Path Simulation of Sports Competition Venue.** In order to verify the effectiveness and superiority of the algorithm in this paper, using the Matlab 7.6 simulation platform, a large number of path optimization experiments were performed on the classic ant colony algorithm and the hybrid artificial potential field-ant colony algorithm in a  $20 \times 20$  grid environment. Taking into account the actual performance and environmental conditions, the optimal path is not the shortest path. Because the robot takes time and energy to turn, the smoothness of the path is also very important. The fewer inflection points and the smaller the corners are, the better the smoothness is. Based on the above description of the characteristics of the ant colony optimization algorithm, we establish an ant colony algorithm suitable for the technical and tactical decision-making of resistance projects to solve the problem of optimization of actual skills and tactics. The process of ant colony algorithm for technical and tactical decision optimization is as follows: (1) Let the number of iterations be  $I$ . The pheromone  $T$  on various technical and tactical decision-making routes is initialized. (2) Randomly

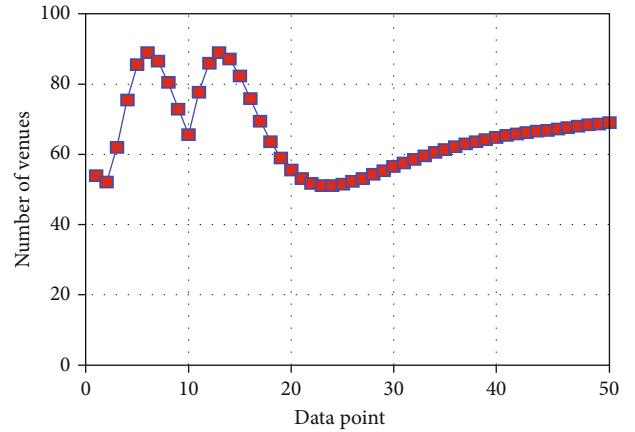


FIGURE 5: Curve of site selection for different data points.

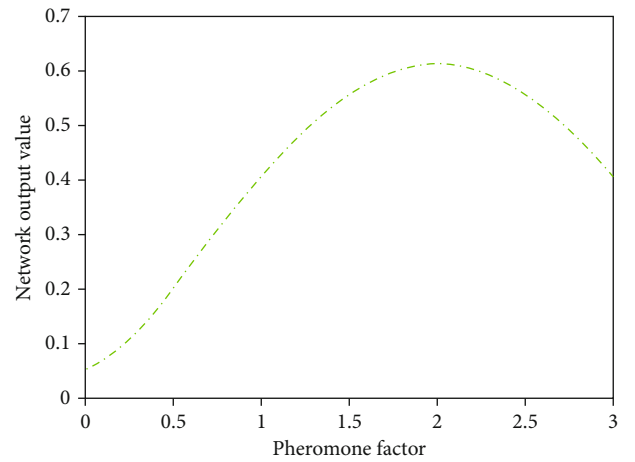


FIGURE 6: Pheromone output value curve of ant colony algorithm.

place  $m$  ants at the end of the technical and tactical decision-making route; establish our technical decision-making taboo table and the opponent's technical decision-making taboo table. (3) Adopt a combination of technical and tactical decision-making and taboo tables to determine the best technical and tactical decision-making route. (4) Construct a solution. Each ant constructs a solution step by step according to the state change rule, that is, a technical and tactical decision route is generated. The task of the ant is to find the best technical decision-making route. Figure 6 shows the pheromone output curve of the ant colony algorithm (see Figure 6).

According to the different mixing ratios of people and vehicles, the ant colony scale is 1000 and running for 200 generations. The two objective function values under different mixing ratios are obtained. It can be seen that the total time of all evacuated objects has decreased first. After the rising trend, it reaches the minimum when the ratio of pedestrians to vehicles is 1:1, and the trend of the total idleness of the road network is still decreasing first and then increasing. The difference is that when the ratio of pedestrians is 80%, the mixed utilization of the road network during the evacuation process is the highest. When pedestrians and

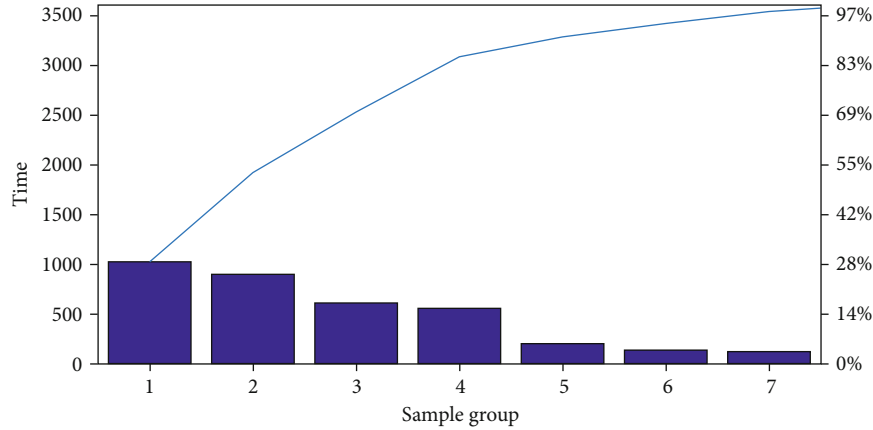


FIGURE 7: Algorithm iteration optimization of different sample groups.

vehicles are mixed in different proportions, the simulation results are shown in paper. According to the above experimental data and analysis, it can be concluded that a single traffic mode cannot get the best evacuation results. When the proportion of pedestrians is 50% to 80%, the effect of mixed evacuation of pedestrians and vehicles is between two values (see Figure 7).

In order to facilitate the comparison of the performance of the two algorithms in path planning, the two algorithms were run 30 times and related indicators were calculated. Considering the optimization results of the aforementioned different mixing ratios in terms of the total evacuation time and the degree of mixed utilization, as well as the evacuation efficiency multiple factors, once again, in mixed evacuation, a high ratio cannot get the best evacuation result. Figure 7 shows the iterative optimization of the algorithm for different sample groups. (1) The obstacle environment coverage rate is 30%, the optimal path is obtained, and the relevant performance indicators of the two algorithms are shown. Experiments show that in a general obstacle environment, both algorithms can converge to the optimal path, but the algorithm in this paper has a faster convergence speed and better optimization effect, and a higher quality of the path is obtained. (2) The barrier environment coverage rate is 35%, and both the optimal path obtained and the related performance indicators of the two algorithms are shown.

**4.3. Example Application and Analysis.** Path planning based on the hybrid artificial potential field-ant colony algorithm is combined with the dual heuristic information of the potential situation and the target distance, the path search is more efficient, the algorithm's global search ability is stronger, and it can effectively avoid falling into the local optimum; at the same time, it adopts the pheromone update mechanism and improves the quality of the path. They are simulations of the evacuation effect of pedestrians and vehicles according to a 1 : 1 ratio of pedestrians and vehicles (500 for pedestrians and vehicles). The evacuation of all objects takes 4660 seconds. With the passage of time, the pheromone left by the previous generation of electronic ants gradually disappears. The parameter is used to indicate the degree of pheromone disappearance. After time, the electronic ant completes a

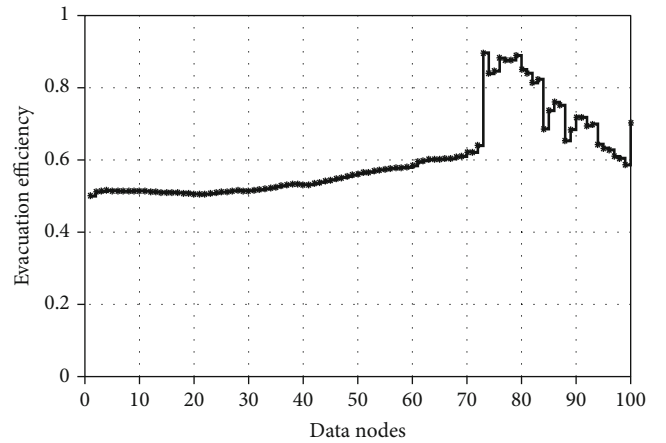


FIGURE 8: The dependence curve of model evacuation efficiency with data points.

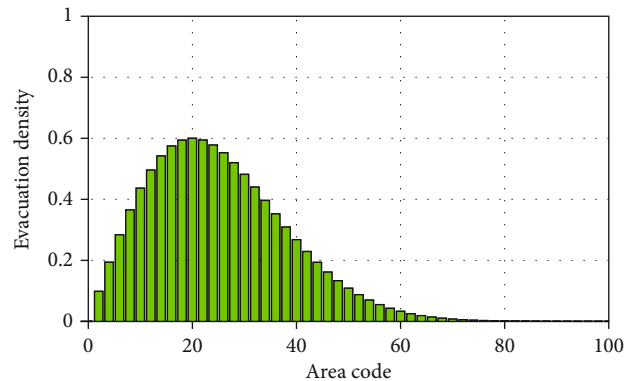


FIGURE 9: Evacuation density distribution map of different subzone codes.

search, and the amount of information on each node is adjusted. It is the distribution of people and vehicles on the road network at the beginning of the evacuation. It can be seen from the figure that as the node goes by, the number of remaining evacuees gradually decreases, and as the evacuation process progresses to 1000 seconds (21% of the total evacuation process), 73.6% of pedestrians have been

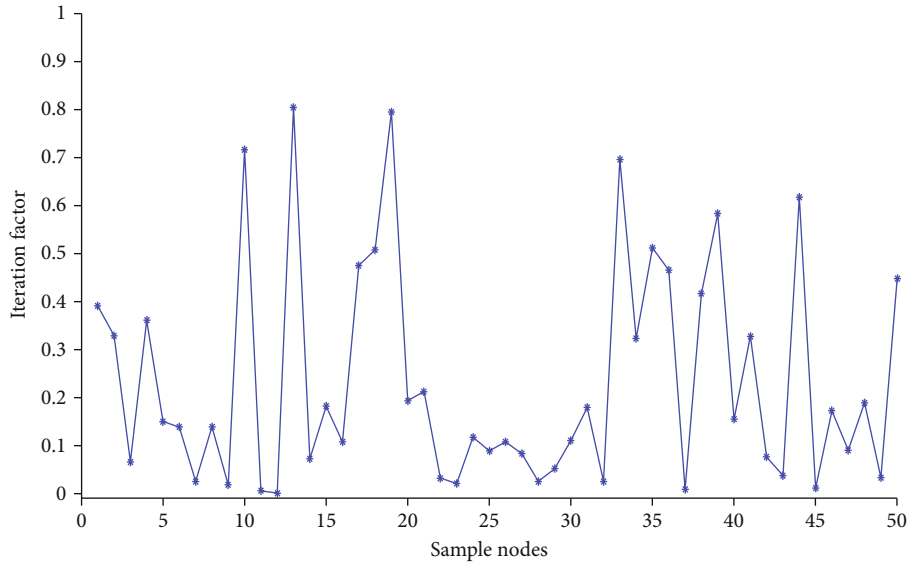


FIGURE 10: Algorithm iteration factor line chart of different nodes.

successfully evacuated and 77.8% of vehicles were evacuated, and at 2000 seconds (43% of the total evacuation process), such as shown in the paper, 95.6% of pedestrians were successfully evacuated and 97% of vehicles evacuated. It can be seen that this model can transfer more than 50% of the evacuees to a safe area within about 20% of the evacuation time. Figure 8 shows the dependence curve of model evacuation efficiency with data points (see Figure 8).

The figure shows the relationship between the percentages of people who have been evacuated over time when the pedestrian mixing ratio ( $R_p$ ) is 10%, 50%, and 90%. It can be seen from the figure that the efficiency of different pedestrian ratios is not different in the later stage of the evacuation process. The main difference lies in the initial stage of the evacuation. The smaller the proportion of pedestrians is, the faster the evacuation is. This is mainly caused by the difference in the movement speed of people and vehicles. A high proportion of vehicles will inevitably speed up the evacuation process. The number of people is gradually mixed with vehicles in the road network. The higher vehicle mixing ratio does not bring the advantage of evacuation efficiency, but reduces it. Figure 9 shows the evacuation density distribution map of different subzone codes. Therefore, we consider the optimization results of the aforementioned different mixing ratios in terms of the total evacuation time and the degree of mixed use of roads, as well as multiple factors of evacuation efficiency; once again, in the mixed evacuation of people and vehicles, a high proportion of pedestrians or vehicles cannot be used. The best evacuation results are obtained; combined with the above analysis, it is concluded that when the pedestrian ratio is minus 80%, the mixed evacuation of pedestrians and vehicles is better in terms of total time, road utilization, and efficiency (see Figure 9).

Experiments show that in a complex environment, the ant colony algorithm has low optimization efficiency, and it is easy to fall into a local optimum. The algorithm in this paper has a stronger optimizing ability, faster convergence speed, and better environmental adaptability. According to

the simulation results, the basic multiobjective ant colony algorithm and the improved multiobjective ant colony algorithm are used to conduct experiments according to the ratio of people to vehicle of 1:1. Obviously, the improvement Pareto frontier obtained by the multiobjective ant colony algorithm is better than the basic ant colony multiobjective optimization algorithm. As time goes by, the evacuated objects are gradually distributed on various road sections. Figure 10 shows the line graph of the algorithm iteration factor for different nodes (see Figure 10).

Comparing the results of the two algorithms, it can be found that there are basically no people stranded in the stadium. In it, there are still people in the stadium that have not been evacuated outside within 600-1000 seconds. The number of individuals in the marked area is significantly lower than that in the corresponding area. The above analysis shows that the improved multiobjective ant colony algorithm is better than the basic multiobjective ant colony algorithm in terms of evacuation efficiency and safety.

## 5. Conclusion

The vector pheromone routing method based on the priority Pareto partial order relationship can give priority to factors closely related to evacuation performance such as evacuation efficiency and congestion conditions (such as the distance to the center of the stadium and the distance to the exit), and effectively filter out secondary factors. The interference of factors can more effectively improve the evacuation efficiency, congestion, and other evacuation performance. In order to improve the congestion of the evacuation plan obtained by the fragmented multiobjective evacuation route optimization algorithm, the vector pheromone routing method based on the priority Pareto partial order relation is used to replace the vector based on the traditional Pareto partial order relation in the algorithm. In this paper, a priority-based Pareto partial order relationship is proposed, and based on this, a vector pheromone routing method based

on the priority Pareto partial order relationship is proposed. Compared with the vector pheromone routing method based on the traditional Pareto partial order relationship, the vector pheromone routing method based on the priority Pareto partial order relationship can give priority to factors closely related to evacuation performance such as evacuation efficiency and congestion, and filter interference of secondary factors can more effectively improve the evacuation efficiency, congestion, and other evacuation performance. Aiming at the mixed evacuation of people and vehicles in the integrated environment of large buildings and road networks, heuristic information, taboo rules, and information update strategies suitable for this problem are proposed. The simulation results show the effectiveness and feasibility of the model and algorithm. Through the analysis of the evacuation performance under different mixture ratios of people and vehicles, the effect of the mixed evacuation of people and vehicles is better. This paper compares the pros and cons of the solution obtained by the basic ant colony algorithm and the improved method, as well as the differences in the temporal and spatial distribution of the evacuated objects. It shows that the improved multiobjective ant colony optimization algorithm can solve the large-scale mixed evacuation problem more quickly and safely. The advantages and disadvantages of the Pareto solution obtained by the basic ant colony algorithm and the improved method are compared, as well as the differences in the temporal and spatial distribution of evacuated objects, which shows that the improved multiobjective ant colony optimization algorithm can solve the large-scale mixed evacuation problem of people and vehicles more quickly and safely.

## Data Availability

All information is within the paper.

## Conflicts of Interest

No competing interests exist concerning this study.

## References

- [1] Y. Wang, "Application of ant colony algorithm and artificial intelligence in training simulation of athletes in sports arena," *Journal of Ambient Intelligence and Humanized Computing*, vol. 3, 2021.
- [2] H. Guangdong, L. Ping, and W. Qun, "A hybrid metaheuristic ACO-GA with an application in sports competition scheduling," in *Eighth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing (SNPD 2007)*, pp. 611–616, Qingdao, China, 2007.
- [3] H. Guangdong and W. Qun, "A hybrid ACO-GA on sports competition scheduling," in *Ant colony optimization- Methods and applications*, pp. 89–100, IntechOpen, 2019.
- [4] R. Seidlova, J. Poživil, and J. Seidl, "Marketing and business intelligence with help of ant colony algorithm," *Journal of Strategic Marketing*, vol. 27, no. 5, pp. 451–463, 2019.
- [5] W. Liu, H. L. Ma, and A. Walsh, "Advance in photonic crystal solar cells," *Renewable and Sustainable Energy Reviews*, vol. 116, 2019.
- [6] X. Zhang, C. Zang, H. L. Ma, and Z. J. Wang, "Study on removing calcium carbonate plug from near wellbore by high-power ultrasonic treatment," *Ultrasonics Sonochemistry*, vol. 62, 2020.
- [7] H. L. Ma, X. Zhang, F. F. Ju, and S. B. Tsai, "A study on curing kinetics of nano-phase modified epoxy resin," *Scientific Reports*, vol. 8, no. 1, p. 3045, 2018.
- [8] M. Ling, M. J. Esfahani, H. Akbari, and A. Foroughi, "Effects of residence time and heating rate on gasification of petroleum residue," *Petroleum Science and Technology*, vol. 34, 2016.
- [9] H. L. Ma and S. B. Tsai, "Design of research on performance of a new iridium coordination compound for the detection of Hg<sup>2+</sup>," *International Journal of Environmental Research and Public Health*, vol. 14, no. 10, p. 1232, 2017.
- [10] L. Y. Mo, W. H. Z. Sun, S. Jiang et al., "Removal of colloidal precipitation plugging with high-power ultrasound," *Ultrasonics Sonochemistry*, vol. 69, p. 105259, 2020.
- [11] D. Gao, Y. Liu, Z. Guo et al., "A study on optimization of CBM water drainage by well-test deconvolution in the early development stage," *Water*, vol. 10, 2018.
- [12] S. B. Tsai and H. Ma, "A research on preparation and application of the monolithic catalyst with interconnecting pore structure," *Scientific Reports*, vol. 8, no. 1, 2018.
- [13] J. Xie and H. Ma, "Application of improved APO algorithm in vulnerability assessment and reconstruction of microgrid," *IOP Conference Series: Earth and Environmental Science*, vol. 108, no. 5, 2018.
- [14] F. Ai, X. Yin, R. Hu, H. Ma, and W. Liu, "Research into the super-absorbent polymers on agricultural water," *Agricultural Water Management*, vol. 245, p. 106513, 2021.
- [15] G. Kendall, S. Knust, C. C. Ribeiro, and S. Urrutia, "Scheduling in sports: an annotated bibliography," *Computers & Operations Research*, vol. 37, no. 1, 2010.
- [16] I. Fister, S. Rauter, X.-S. Yang, K. Ljubič, and I. Fister, "Planning the sports training sessions with the bat algorithm," *Neurocomputing*, vol. 149, pp. 993–1002, 2019.
- [17] D. Karaboga, K. Guney, and A. Akdagli, "Antenna array pattern nulling by controlling both amplitude and phase using modified touring ant colony optimization algorithm," *International Journal of Electronics*, vol. 91, no. 4, pp. 241–251, 2004.
- [18] M. R. Zeng, L. Xi, and A. M. Xiao, "The free step length ant colony algorithm in mobile robot path planning," *Advanced Robotics*, vol. 30, no. 23, pp. 1509–1514, 2018.
- [19] H. C. Huang, "Intelligent motion control for omnidirectional mobile robots using ant colony optimization," *Applied Artificial Intelligence*, vol. 27, no. 3, pp. 151–169, 2013.
- [20] A. Boulmakoul and M. Mandar, "Fuzzy ant colony paradigm for virtual pedestrian simulation," *The Open Operational Research Journal*, vol. 5, no. 1, pp. 1–8, 2019.
- [21] X. Chen, L. Yu, T. Wang et al., "Artificial intelligence-empowered path selection: a survey of ant colony optimization for static and mobile sensor networks," *IEEE Access*, vol. 8, pp. 71497–71511, 2020.
- [22] R. S. Parpinelli, H. S. Lopes, and A. A. Freitas, "Data mining with an ant colony optimization algorithm," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 4, pp. 321–332, 2002.

- [23] Y. Zhang, Y. Jiang, L. Qi, M. A. Bhuiyan, and P. Qian, "Epilepsy diagnosis using multi-view & multi-medoid entropy-based clustering with privacy protection," *ACM Transactions on Internet Technology*, vol. 21, no. 2, pp. 1–21, 2021.
- [24] Y. Zhang, G. Wang, F. L. Chung, and S. Wang, "Support vector machines with the known feature-evolution priors," *Knowledge-Based Systems*, vol. 223, 2021.
- [25] Y. Zhang, S. Wang, K. Xia, Y. Jiang, and P. Qian, "Alzheimer's disease multiclass diagnosis via multimodal neuroimaging embedding feature selection and fusion," *Information Fusion*, vol. 66, pp. 170–183, 2021.
- [26] T. Grubljesic, P. S. Coelho, and J. Jaklic, "The shift to socio-organizational drivers of business intelligence and analytics acceptance," *Journal of Organizational and End User Computing*, vol. 31, no. 2, pp. 37–64, 2019.
- [27] L. X. Z. Zhang, M. Mouritsen, and J. R. Miller, "Role of perceived value in acceptance of "bring your own device" policy," *Journal of Organizational and End User Computing*, vol. 31, no. 2, pp. 65–82, 2019.
- [28] A. Shahri, M. Hosseini, K. Phalp, J. Taylor, and R. Ali, "How to engineer gamification," *Journal of Organizational and End User Computing*, vol. 31, no. 1, pp. 39–60, 2019.

## Research Article

# A College Student Behavior Analysis and Management Method Based on Machine Learning Technology

Xiaoying Shen <sup>1</sup> and Chao Yuan <sup>2,3</sup>

<sup>1</sup>Wuxi Vocational College of Science and Technology, No. 8 Xinxi Road, Wuxi, Jiangsu 214000, China

<sup>2</sup>School of Design, Jiangnan University, 1800 Lihu Avenue, Wuxi, Jiangsu 214122, China

<sup>3</sup>College of Economics and Management, Nanjing University of Aeronautics and Astronautics, 29 Jiangjun Avenue, Nanjing, 211100 Nanjing, China

Correspondence should be addressed to Chao Yuan; [circle@jiangnan.edu.cn](mailto:circle@jiangnan.edu.cn)

Received 30 June 2021; Revised 16 July 2021; Accepted 20 July 2021; Published 2 August 2021

Academic Editor: Shan Zhong

Copyright © 2021 Xiaoying Shen and Chao Yuan. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A digital campus will generate a large amount of student-related data. How to analyze and apply these data has become the key to improving the management level of students. The analysis of student behavior data can not only assist schools in early warning of dangerous events and strengthen school safety but also can use real data to describe student behavior, thereby providing quantitative data support for scholarship and grant evaluation. This paper takes a university student as the research object, collects various data in the digital campus platform, and uses an adaptive *K*-means algorithm in the machine learning algorithm to cluster the data. Analyze the behavior of college students from the clustering results, so as to provide a basis for the education management and learning ability improvement of college students. Specifically, the student's study, life, and consumption data are selected as the data to describe the student's behavior at school. This data is input into the adaptive *K*-means algorithm to obtain different types of student consumption habits, living habits, and learning habits. Through the analysis results, it can be found that the problem of the group of students with low financial ability, the problem of too long online time for students, and the number of books borrowed are too low. According to the characteristics of these problems, teachers and schools are provided with targeted management suggestions. The analysis of student behavior based on machine learning technology provides a reference for the formulation of students' school management policies and provides teachers with information on students' personality characteristics, which is conducive to improving teachers' teaching effects. In short, the management of the results of student behavior analysis can provide a basis for the school to formulate reasonable management policies, thereby promoting precision management and scientific decision-making.

## 1. Introduction

The establishment of a digital campus has improved the efficiency of university management and has also brought great convenience to students, faculty, and staff. The digital management system can collect a large amount of data, which plays an important role in the management of the school. As for the daily management of students, if we can learn more about students, we can implement more effective programs for different students, so that we can teach students in accordance with their aptitude and improve the education level of the school. The traditional analysis and management

of student behavior mostly relies on the personal experience of the manager and lacks the individualized cognition of the learner. At the same time, it cannot in-depth guide students' learning behaviors, provide personalized learning situations, and promote learning optimization. Analyzing student life and learning behavior based on intelligent technology are of great significance to the investigation of potential abnormal students and the prediction of students' future development. The key to understanding students is the data collected in the digital campus for students' study, life, and consumption. At present, many schools have established corresponding all-in-one card systems, which make students'

daily campus life more convenient. Students can use the campus card to consume in canteens, supermarkets, etc., or use the card to borrow books in the library, etc. These operations will generate a large amount of student behavior data. How to use these data to discover the information contained in it is a problem that needs to be solved. Based on the machine learning technology, this paper conducts cluster analysis on campus all-in-one card data and analyzes the behavior of students.

In recent years, there has been a lot of research on student behavior analysis. Reference [1] measures student behavior based on entropy measurement. The study defined two behavioral characteristics of orderliness and diligence and analyzed the correlation between the regularity of campus life and academic performance. Reference [2] proposed a pre-class student performance prediction method based on multiexample multilabel learning. The idea of this method is to use students' behavior in completed courses to predict their difficulties in learning new courses. The results of this study are convenient for teachers to track and understand the learning situation of each student. Reference [3] proposes an education measurement system to characterize educational behavior by collecting campus Wi-Fi network data. The results show that the system can obtain information about the relationship between punctuality, distraction, and academic performance. Reference [4] uses an improved recurrent neural network to simulate the student's answering process according to the student's answer records and the content of each exercise to predict the student's future performance. Based on MOOCs learner behavior data, reference [5] established a prediction model based on clustering algorithm and neural network to mine the learning rules in the learning process. The predicted results can provide personalized guidance for each learner. Reference [6] proposes a classification system to analyze the behavior of students in the teaching system and to find students with poor performance early. Reference [7] predicts their course performance based on the relevant data generated by students during the online course learning process. Reference [8] input multimodel data and semester information into the linear mixed effects model to predict the future performance of students. Reference [9] found that an improved random forest method can be used to predict the grades of freshmen and existing courses. The application of these technologies brings hope to student degree planning, lecturer intervention, and personalized advice. Annapol State University in India has developed a product [10, 11], which is used to monitor student activity areas. This product analyzes students' participation in organizing activities based on the records of students swiping their ID cards. Through this software, information about students who participate in activities with low frequency can be detected.

Most of the above studies have completed the analysis of student behavior based on machine learning technologies [12–14]. The current popular machine learning technologies contain the tradition machine learning methods [15, 16] and some advanced machine leaning technologies [17–19]. These advanced machine leaning technologies have been used in many practical applications, such as medicine [20], industry

[21], and basic theory research [22–24]. In terms of performance, the accuracy of behavior analysis based on deep learning is indeed higher, but its computational complexity and hardware performance requirements are higher. The time complexity of behavior analysis based on machine learning is relatively low, and the algorithm implementation is simple. Therefore, this article chooses student behavior analysis based on machine learning. In machine learning,  $K$ -means clustering algorithm [25, 26] is widely used in student behavior analysis research. However, traditional  $K$ -means is very sensitive to outliers, and a small number of outliers will have a great impact on the final clustering results. On the other hand, the  $K$ -means algorithm still has the problem that the  $K$  value cannot be adapted. In response to these problems, this paper uses an adaptive  $K$ -means algorithm to improve the efficiency and clustering accuracy of the algorithm.

The main work of this paper is as follows:

- (1) Collect student consumption, life, and learning data through the campus all-in-one card system and integrate data from different institutions to form a comprehensive data set for student behavior analysis
- (2) An adaptive  $K$ -means clustering algorithm is used for student behavior analysis. The algorithm introduces the elbow rule to optimize the data and find points far away from the cluster, thereby effectively improving the clustering performance. In addition, the algorithm also introduces the idea of self-adaptation and automatically adjusts the value of  $k$  based on the sum of squared errors. The adaptive  $K$  value is more suitable for objective reality
- (3) Based on the above model, the analysis results of students' consumption, life, and learning are obtained. The analysis results can be used to improve the management level of the school and truly teach students in accordance with their aptitude

## 2. Related Work

*2.1. Student Behavior Analysis and Management.* The complexity of individual college students makes it impossible for school administrators to understand students' dynamics in real time. For some students with abnormal behaviors, the students around them may be inconvenient or embarrassed to inform the management staff of the specific situation, which causes a certain lag in the work of the student management staff. In order to understand the behavior and habits of students in real time, the management of students will be transformed from passive to active. The daily behavior data of students needs to be displayed intuitively on the student behavior analysis system. On the premise of ensuring the privacy and safety of students, the data of the system mainly comes from the data collected by the school's digital systems, and the machine learning technology is used to analyze the data. The goal of student behavior analysis is shown in Figure 1.



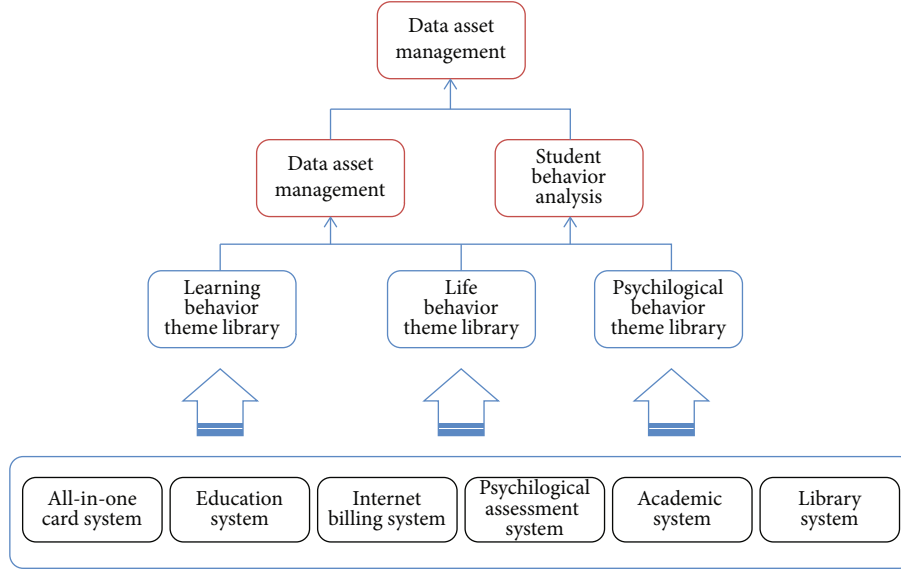


FIGURE 1: Student behavior analysis and management goals.

As shown in Figure 1, first, the historical data stored in the school's digital management systems need to be integrated and stored. Second, establish a data analysis model based on the dimensions of student behavior data analysis. Third, establish a student behavior data analysis system to achieve student management goals based on data services. Finally, a real-time monitoring system for student behavior is established to monitor abnormal students in real time.

**2.2. Student Behavior Analysis Process Based on Machine Learning.** The process of applying machine learning to student behavior analysis is shown in Figure 2. As shown in Figure 2, it is first necessary to collect and preprocess student behavior data. Collection is by exporting system data such as the school's teaching system and campus card. The format and structure of the exported data vary greatly. Various data needs to be integrated to obtain comprehensive data. Second, because there will be a lot of noise data in the integrated data, the data needs to be preprocessed. The preprocessed data often has problems such as high dimensionality, so it is necessary to perform feature extraction on the data. Third, conduct behavior analysis model training based on the training set. Fourth, input the test data into the trained analysis model to obtain the analysis result. Finally, perform related management and application based on the analysis results.

### 3. Student Behavior Analysis Based on Adaptive K-Means

**3.1. Behavior Analysis Framework.** As shown in Figure 3, first, integrate the data collected by the campus card, educational administration system, etc. These data are mainly composed of students' life, consumption, and learning data. Second, because there is noisy data in the integrated data, it is necessary to perform preprocessing such as cleaning the data. Third, perform feature extraction on the preprocessed data and extract the main features for subsequent processing.

The feature extraction method used in this paper is principal component analysis (PCA) [27, 28]. Fourth, divide the feature data set into a training set and a test set. The training set is used to train the behavior analysis model. Fifth, the test set is input to the analysis model to obtain the analysis result. Finally, the analysis results are managed and applied.

**3.2. Behavior Analysis Model.** The analysis model used in Figure 3 is an adaptive K-means algorithm. The idea of the algorithm used is to optimize the sample data set  $X$  through the elbow rule to determine outliers. When the algorithm is implemented, the sample data set that eliminates outliers is used, and after the algorithm is completed, the final outlier is determined according to the similarity between the outliers and each cluster. Based on the adaptive idea, after each iteration is completed, the value of  $k$  is automatically adjusted according to the cluster evaluation index error of each cluster until the error range is met.

**3.2.1. The Elbow Rule Detects Outliers.** The similarity determination of traditional K-means algorithm is based on Euclidean distance. Outliers will affect the estimation of  $k$  value, thereby increasing the time complexity of the algorithm. Use the elbow method to effectively detect outliers in the data set to optimize the algorithm. The specific implementation is as follows:

Let the data set be  $X = \{x_i | i = 1, 2, \dots, m\}$  and  $m$  be the number of samples. Each sample has  $n$  features ( $n > 0$ ); the sample is divided into different categories  $C = \{c_1, c_2, \dots, c_k\}$ ,  $k$  is the number of clusters. Initially, all samples in  $X$  are regarded as one class, and the initial class center is

$$\mu_j = \frac{1}{N(C_j)} \sum_{x_i \in C_j} x_i, \quad (1)$$

where  $C$  represents the sample set contained in the  $j$ -th cluster and  $N(C_j)$  represents the number of samples in the

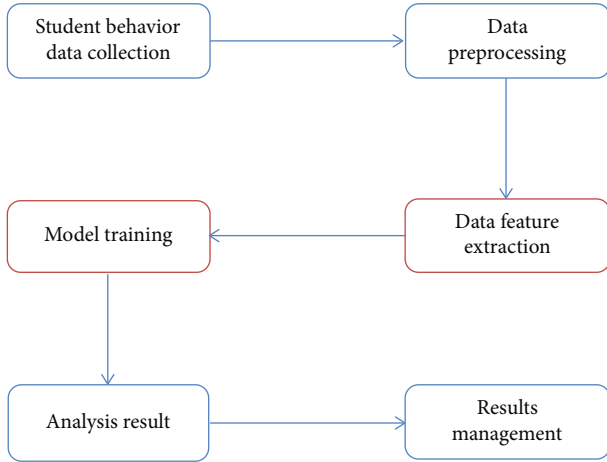


FIGURE 2: Flow chart of student behavior analysis based on machine learning.

$j$ -th cluster. Calculate the Euclidean distance set  $D$  from each sample in  $X$  to the cluster center  $\mu_j$ ; the specific formula is as follows:

$$d(x_i, \mu_j) = \sqrt{(x_i - \mu_j)^2}, \quad (2)$$

where  $x_i$  represents the  $i$ -th sample,  $i \in [1, m]$ ,  $D = \{d_1, d_2, \dots, d_m\}$ .

According to the elbow rule, sort the data in  $D$  from small to large and get an  $x$ - $d$  two-dimensional line graph, where  $d$  is the distance from the sample data point to the center point and  $x$  is the sample corresponding to  $d$ . As  $d$  increases, the value corresponding to the position where the distortion improvement effect increases the most is the elbow. Therefore, the elbow meets the following conditions:

$$\Delta d = \max(d_i - d_{i-1}), i \in [1, m]. \quad (3)$$

Divide at the elbow and temporarily define data whose distance is greater than the corresponding distance of the elbow point as an outlier. Assuming there are  $w$  samples ( $w \leq m$ ) after eliminating outliers, there are  $(m-w)$  outliers. Store the nonoutliers in the sample set  $X'$  and renumber the samples to get  $X' = \{x_1, x_2, \dots, x_w\}$ . The outliers are stored in the data set  $Y$  and renumbered to obtain  $Y' = \{x_{w+1}, x_{w+2}, \dots, x_m\}$ . The use of  $X'$  sample set in the implementation of the algorithm can eliminate the influence of outliers to a certain extent.

After the algorithm is implemented, cluster  $C = \{c_1, c_2, \dots, c_z\}$  is obtained,  $z$  is the number of clusters, and the maximum distance from the sample in the cluster to the center of the cluster in the  $j$ -th cluster is  $\max d_j$ .

$$\max d_j = \max \{d_{j1}, d_{j2}, \dots\}. \quad (4)$$

According to Equation (2), calculate the distance  $\{(d_{(w+1)1}, d_{(w+1)2}, \dots, d_{(w+1)z}), (d_{(w+2)1}, d_{(w+2)2}, \dots, d_{(w+2)z}), \dots,$

$(d_{m1}, d_{m2}, \dots, d_{mz})\}$  from the sample in  $Y$  to the center of each cluster. If there is a cluster  $(a \in (0, z))$ , make the sample  $x_b (b \in [w+1, m])$  in the set  $Y$  satisfy that the distance from  $x_b$  to the center of cluster  $a$  is less than the maximum distance from the samples in the cluster to the center of the cluster, namely,  $(d_{b1} < \max d_1) \parallel (d_{b2} < \max d_2) \parallel \dots \parallel (d_{ba} < \max d_a)$ . Then, divide  $x_b$  into the nearest cluster among  $a$  clusters.

$$j(x_b) = \min(d_{b1}, d_{b2}, \dots, d_{ba}). \quad (5)$$

If there is no such cluster, the sample is defined as an outlier. Until all the samples in  $Y$  are traversed, the final outliers can be divided.

3.2.2. *Selection of Adaptive  $k$  Value.* Appropriate selection of  $k$  value needs to be evaluated based on clustering evaluation index. This article uses the sum of squares of errors SSE within the cluster. The calculation formula of this indicator is as follows:

$$E = \sum_{j=1}^k \sum_{x \in C_j} |x - \mu_j|^2, \quad (6)$$

where  $k$  represents the number of clusters,  $x$  represents samples,  $\mu_j$  represents the cluster center of the  $j$ -th cluster, and  $C_j$  represents the set of samples contained in the  $j$ -th cluster.  $E$  describes the tightness of each cluster sample to a certain extent; the smaller the  $E$ , the better the clustering effect. According to the SSE, the sum of squared errors  $Je$  in each cluster is obtained. The calculation formula of  $Je$  is as follows:

$$Je_j = \sum_{x \in C_j} \frac{1}{N(C_j)} |x_i - \mu_j|^2, \quad (7)$$

where  $x_i$  is the sample in the  $j$ -th cluster,  $N(C_j)$  is the number of samples in the  $j$ -th cluster, and  $\mu_j$  is the sample in the  $j$ -th cluster. The smaller  $Je_j$  means the better the clustering effect of the  $j$ -th cluster.

Initially, set  $Je$  and the threshold  $N$  of the minimum number of samples in the cluster. After each cluster is divided, the number of samples  $(N_1, N_2, \dots, N_k)$  in each cluster can be obtained, and the  $(Je_1, Je_2, \dots, Je_k)$  of each cluster can be calculated according to Equation (7). Then, calculate the error  $\Delta Je_j$  of the  $j$ -th cluster clustering evaluation index and the difference  $\Delta N_j$  between the number of samples in the cluster and the initial value. The specific formula is as follows:

$$\Delta Je_j = Je_j - b, \quad (8)$$

$$\Delta N_j = N_j - N. \quad (9)$$

Combine Equation (8) and Equation (9) to get the change

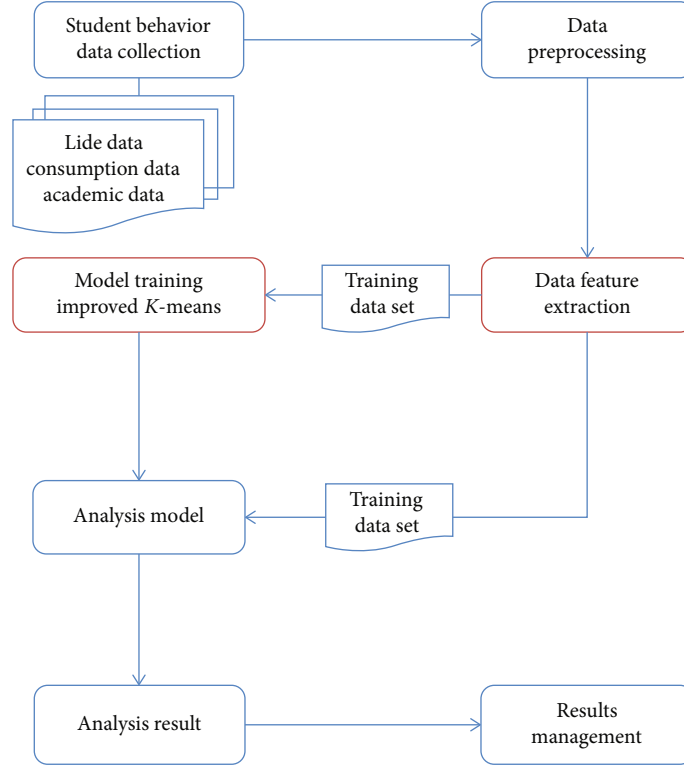


FIGURE 3: Architecture diagram of student behavior analysis.

of  $k$  value in the  $j$ -th cluster

$$\Delta k_j = \frac{\text{sgn}(\Delta N_j) - 1}{2} + \frac{\text{sgn}(\Delta N_j) + 1}{2} * \Pi(\log_w(\Delta J e_j + 1))\theta(\Delta J e_j), \quad (10)$$

where  $w$  is the number of samples in the data set  $X'$ ,  $\text{sgn}()$  is the symbolic function,  $\theta()$  is the unit step function, and the symbol  $\Pi()$  is rounded up.

If  $N_j < N$ , then  $\Delta N_j$  is negative,  $\text{sgn}(\Delta N_j) = -1$ , so  $(\text{sgn}(\Delta N_j) + 1)/2 = 0$ ,  $(\text{sgn}(\Delta N_j) - 1)/2 = -1$ .  $\Delta k_j = -1$  represents that when the number of samples in the  $j$ -th cluster is less than the initial value, delete the cluster center of the cluster.

If  $N_j > N$ , then  $\Delta N_j$  is a positive number,  $\text{sgn}(\Delta N_j) = 1$ ,  $(\text{sgn}(\Delta N_j) + 1)/2 = 1$ ,  $(\text{sgn}(\Delta N_j) - 1)/2 = 0$ . The discussion is divided into the following two situations:

- (1) When  $J e_j > b$ , then  $J e_j$  is a positive number,  $\theta(J e_j) = 0$ , and  $\Delta k_j = \Pi(\log_m(\Delta J e_j + 1))$ , so  $\Delta k_j > 0$ . It is necessary to add a new cluster center near the  $j$ -th cluster center to reduce the error evaluation index within the cluster. Generally,  $0 < \log_m(\Delta J e_j + 1) < 1$ , then  $\Delta k_j = 1$ . Only when the error  $\Delta(J e_j)$  is particularly large,  $\log_m(\Delta J e_j + 1)$  will be greater than 1
- (2) When  $J e_j < b$ , then  $J e_j$  is negative,  $\theta(J e_j) = 0$ , and  $\Delta k_j = 1$ .  $\Delta k_j > 0$  requires a new cluster center near the  $j$ -th cluster center to reduce the error evaluation

index within the cluster. Generally,  $0 < \log_m(\Delta J e_j + 1) < 1$ , then  $\Delta k_j = 0$ . When the cluster evaluation index in the  $j$ -th cluster is less than the set initial value, the cluster center of the cluster is neither deleted nor added. The sample closest to the  $j$ -th cluster center is set as the new cluster center to reduce the clustering evaluation index within the cluster. After traversing each cluster, according to Equation (10), the amount of change in  $k$  value  $(\Delta k_1, \Delta k_2, \dots, \Delta k_k)$  is obtained. The updated  $k'$  is

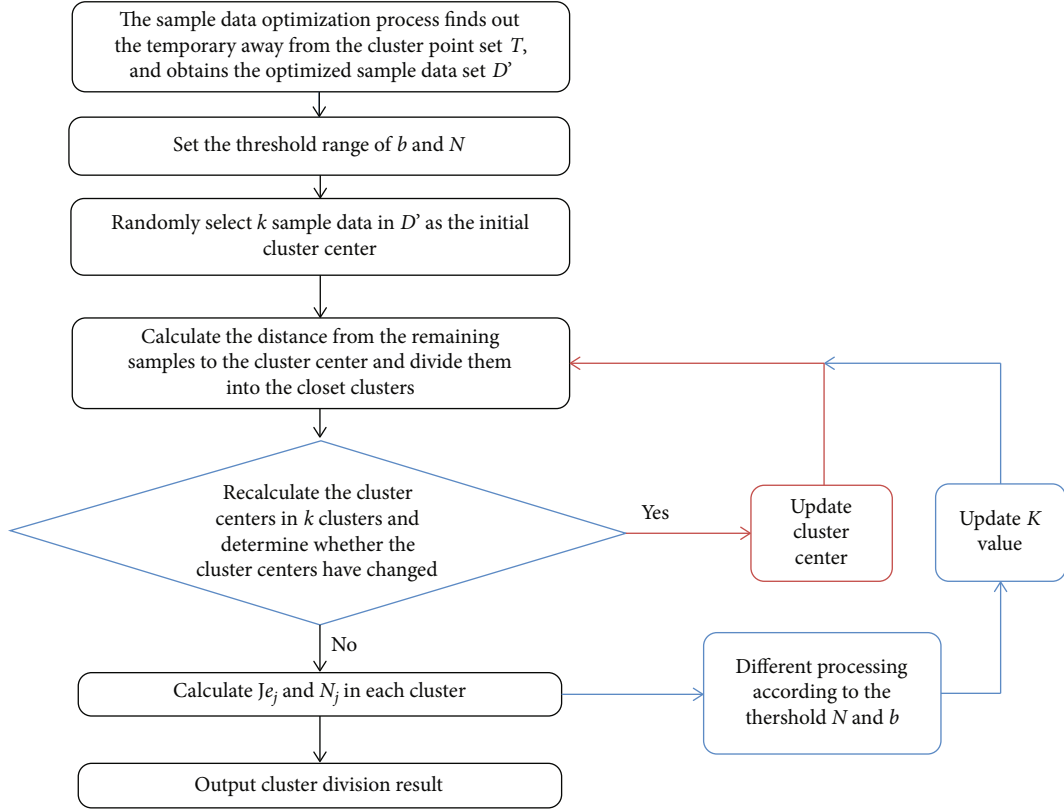
$$k' = k + \sum_{j=1}^k \Delta k_j. \quad (11)$$

If  $k' = k$ , terminate the loop; if  $k' \neq k$ , continue the loop. The flow of the algorithm is shown in Figure 4.

The specific implementation steps of the algorithm are as follows:

*Step 1.* Data optimization processing. Find outliers according to the elbow rule and store them in sample set  $Y$ . The new sample data set after optimization is  $X' = \{x_1, x_2, \dots, x_w\}$ .

*Step 2.* Set the threshold range of the evaluation index  $J e$  in a single cluster and the minimum number of samples  $N$  in the cluster.

FIGURE 4: Adaptive  $K$ -means algorithm flow.

*Step 3.* Randomly select  $k$  samples from  $w$  samples as the initial cluster center  $1 < k \leq w$ .

*Step 4.* Calculate the Euclidean distance of each remaining sample to the cluster center according to Equation (2) and divide each sample into the cluster closest to it.

*Step 5.* Recalculate the new cluster center of each cluster according to Equation (1).

*Step 6.* If the new cluster center is the same as the original center or less than a certain threshold, the iteration is terminated. If the new cluster center changes, continue to repeat Steps 4 and 5 until convergence.

*Step 7.* Calculate the  $Je_j$  of each cluster and the number of samples  $N$  in each cluster. Compare  $Je_j$  and  $N_j$  with the initial threshold range and calculate  $\Delta k_j$  according to Equation (10). After traversing all the clusters, calculate the new  $k$  value  $k'$  according to Equation (11). If any center is deleted or added, return to Step 4 until there are no new or deleted cluster centers.

*Step 8.* Determine the final divided outliers according to the similarity between the samples in the sample set  $Y$  and the cluster centers of each cluster.

*Step 9.* Output the final cluster partition  $C = \{c_1, c_2, \dots, c_z\}$ .

## 4. Experiment and Analysis

*4.1. Experimental Background.* The data used in this article comes from data in a university's digital system database. The consumption data comes from the campus all-in-one card system, which mainly includes canteen consumption and school supermarket consumption. The life data comes from the all-in-one card system, which mainly includes exercise clock-in and time spent online. The learning data comes from the educational administration system and the book borrowing system. A total of 2017 students were selected as a sample. The extracted raw data mainly contains 89,682 pieces of consumption data, 49,860 pieces of life data, and 15,629 pieces of learning data. After the integration and preprocessing of the data, sample data of 3356 students were obtained as experimental data. Table 1 is the definition of students' consumption habits, Table 2 is the definition of students' living habits, and Table 3 is the definition of students' learning habits.

The hardware configuration information used in this experiment is as follows: CPU is Intel Core i7, graphics memory is GTX960M 4G, and memory is 16G. The operating system is Windows 10 64-bit, and the development language is MATLAB.

### 4.2. Experimental Results and Analysis

*4.2.1. Analysis of Consumption Data.* The results of cluster analysis of consumption data based on the algorithm used in this article are shown in Table 4.

TABLE 1: Student consumption habit indicators.

Index	Ranges	Index description
Average monthly consumption	0-2000	The total consumption per student per semester divided by the number of semester months
Average monthly consumption frequency	0-500	The sum of consumption times per student per semester divided by the number of semester months
Peak monthly consumption	0-unknown	Peak consumption of each student in the months of each semester

TABLE 2: Student living habit indicators.

Index	Ranges	Index description
Dining habits	0-90	The average number of days the student eats regularly per month in the semester
Work and rest	0-30	The average number of days of regular work and rest per month for students in each semester
Internet habits	0-600	Average online time per month per semester by students
Exercise habits	0-60	The average number of student exercises per month per semester

TABLE 3: Student study habit indicators.

Index	Ranges	Index description
Attend class	0-1	Number of student attendance in class
Book borrowing	0-unknown	Number of books borrowed by students per semester
	0-unknown	Number of times the student enters the library per semester

TABLE 4: Cluster analysis results of student consumption data.

Number	Student ratio	Monthly consumption	Monthly consumption	Peak monthly consumption
1	10.78	150.23	586.56	652.14
2	25.16	136.52	1091.82	1356.60
3	34.51	112.10	829.75	928.03
4	16.45	77.79	678.50	788.50
5	13.10	45.97	467.92	600.31
Mean		104.52	730.91	865.12

It can be seen from Table 4 that the most suitable  $k$  value obtained based on the  $K$ -means clustering algorithm used in this article is 5, which shows that 5 types of consumption habits can be obtained according to the consumption data of students. Each habit corresponds to different types of students. The characteristics of each type of student are as follows:

- (1) Type 1 students have lower monthly consumption levels and single-month peak consumption, but they consume more times. This type belongs to the group with lower consumption levels. Such students have poor family economic conditions and live frugal lives. It is recommended that school administrators pay attention to the living conditions of such students, and the identification and funding of poor students can consider choosing from such students
- (2) The average monthly consumption of type 2 students is the highest, the number of consumption is high,

and the consumption peak is also high, indicating that this type of student is a high consumption group in the cafeteria

- (3) The monthly consumption level of type 3 students is above average, and the consumption frequency is higher each month, and the consumption amount is stable. It shows that such students often eat in the school cafeteria, and their consumption is stable, which is in line with the normal eating rules of most students in school
- (4) The monthly consumption level of type 4 students is in the middle, and the average monthly consumption is not high. However, the maximum consumption in a single month is relatively high, and the number of consumptions is also relatively small. Such students eat irregularly in the cafeteria and usually like to eat out of school or order takeaways
- (5) The average monthly consumption of type 5 students is relatively low, the number of times is relatively small, and the maximum consumption is not high. Such students do not consume frequently in the cafeteria and are more likely to eat outside of school and consume more outside of school. School administrators should pay attention to the food safety and personal safety of such students

## 5. Life Data Analysis

Table 5 shows the results of clustering analysis of life data based on the algorithm used in this paper.

TABLE 5: Cluster analysis results of student life data.

Number	Student ratio	Dining	Work and rest	Internet	Exercise
1	38.82	79.80	21.02	381.66	45.93
2	9.86	28.67	7.16	548.50	6.93
3	51.32	40.48	18.85	420.24	16.40
Mean		49.65	15.68	450.13	23.09

TABLE 6: Cluster analysis results of student learning data.

Number	Student ratio	Class attendance	Number of books borrowed	Number of visits to the library
1	11.98	99.58	10.37	24.75
2	53.36	94.20	7.80	21.44
3	24.02	90.08	8.22	15.03
4	10.64	96.57	5.67	10.47
Mean		95.11	8.02	17.92

For the cluster analysis of life data, three categories were gathered. According to the information in Table 5, the following inference can be drawn.

- (1) Type 1 students have regular schedules and meals every month. They spend a long time online and often participate in physical exercises. Such students have strong self-discipline, good physical fitness, and good living habits
- (2) Type 2 students overslept more frequently each month. They eat irregularly in the cafeteria, spend a long time online, and exercise less frequently. Such students have poor physical fitness. School administrators should pay attention to the learning and class conditions of such students and whether they often skip classes
- (3) Type 3 students often get up early every month, but they have irregular meals in the cafeteria, spend more time online, and do less physical exercise. Such students should not have a healthy habit of eating breakfast and do not like to exercise. School administrators should urge such students to change their unhealthy living habits and pay attention to their health

## 6. Learning Data Analysis

The results of cluster analysis of academic data based on the algorithm used in this paper are shown in Table 6.

Based on the *K*-means algorithm used in this article to analyze the learning data, students are divided into 4 categories. Based on the information shown in Table 6, the characteristics of each type of student are as follows:

- (1) Type 1 students' classroom attendance, the number of books borrowed in the library, and the number

of library entrances are all high. Such students study hard and have good study habits

- (2) Type 2 students have a higher class attendance rate, but the number of books borrowed is not many, and they enter the library more often. Such students are active in class and often read and study in the study room and library. However, the small amount of books they borrow in the library indicates that such students are accustomed to reading books in the library
- (3) Type 3 students have a low class attendance rate, a small amount of books borrowed, and a small number of in and out of the library. It shows that such students often skip classes and do not study hard enough. They are students who do not like to learn. School administrators should focus on the learning situation of such students and promptly urge them to form good learning habits
- (4) Type 4 students have an average class attendance rate, fewer books borrowed, and fewer trips to the library. This kind of students just go to class often, and they do not have much time to study after class, and the degree of study hard is not high. This type of overstudy is a type that does not pay much attention to study at ordinary times and makes surprise review before the exam. It is recommended that such students develop regular study habits and arrange their study time reasonably and appropriately

## 7. Conclusion

In order to assist the school to improve the management level of students, this paper uses an adaptive *K*-means clustering algorithm to analyze the behavior characteristics of students. By collecting the digital system data of the campus, students' consumption, life, and learning data can be obtained. After the data is preprocessed, PCA is used for feature extraction to obtain feature data and perform model training. Input the test set into the trained model to get the clustering result. Finally, based on the analysis of the clustering results, the characteristics of all kinds of students are obtained. According to the results of consumption data analysis, five types of consumer groups were obtained. For groups with low consumption levels, more consideration can be given to poor student evaluation and work-study programs. According to the analysis results of life data, three groups are obtained. Teachers should pay special attention to groups that eat irregularly, spend a long time online, and do not exercise regularly. According to the analysis results of the learning data, 4 groups of groups are obtained. Teachers need to pay more attention to students with average attendance rate, low book reading volume, and small number of library entrances. In the management process of the school, students with different characteristics can be taught in accordance with their aptitude, thereby improving the management quality of the school. There are still some shortcomings in this article; for example, there are some limitations in the description of

student behavior characteristics. In the analysis of student behavior characteristics, due to the limited data sources in the digital campus platform, it is not possible to fully reflect the behavior characteristics of students in school. Later, as the application of the school's digital campus is perfected, more campus businesses will be transferred from traditional offline to online, and more comprehensive student data can be collected.

## Data Availability

The labeled dataset used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare no conflicts of interest.

## Acknowledgments

This work was supported in part by the National Social Science Fund Project under Grant 17BGL102, the Excellent Project of Jiangsu Province Social Science Union under Grant 15SYC-043, the Soft Science Research of Wuxi Science and Technology Association under Grant KX15-B-01, and the Fundamental Research Funds for the Central Universities under Grant 2015ZX18.

## References

- [1] Y. Cao, J. Gao, D. Lian et al., "Orderness predicts academic performance: behavioral analysis on campus lifestyle," *Journal of the Royal Society Interface*, vol. 15, no. 146, article 20180210, 2017.
- [2] Y. Ma, C. Cui, X. Nie, G. Yang, K. Shaheed, and Y. Yin, "Pre-course student performance prediction with multi-instance multi-label learning," *Science China Information Sciences*, vol. 62, no. 2, pp. 200–205, 2019.
- [3] M. Zhou, M. Ma, Y. Zhang, K. SuiA, D. Pei, and T. Moscibroda, "EDUM: classroom education measurements via large-scale WiFi networks," in *ACM International Joint Conference on Pervasive & Ubiquitous Computing*, Heidelberg, Germany, 2016.
- [4] Y. Su, Q. Liu, Q. Liu et al., "Exercise-enhanced sequential modeling for student performance prediction," in *Thirty-Second AAAI Conference on Artificial Intelligence*, New Orleans, Louisiana, USA, 2018.
- [5] Y. Zhang and W. Jiang, "Score prediction model of MOOCs learners based on neural network," *International Journal of Emerging Technologies in Learning (IJET)*, vol. 13, no. 10, pp. 171–182, 2018.
- [6] K. Casey and D. Azcona, "Utilizing student activity patterns to predict performance," *International Journal of Educational Technology in Higher Education*, vol. 14, no. 1, p. 4, 2017.
- [7] J. W. You, "Identifying significant indicators using LMS data to predict course achievement in online learning," *The Internet and Higher Education*, vol. 29, pp. 23–30, 2016.
- [8] D. D. Mitri, M. Scheffel, H. Drachslar, D. Börner, S. Ternier, and M. Specht, "Learning pulse: a machine learning approach for predicting performance in self-regulated learning using multimodal data," in *The Seventh International Learning Analytics & Knowledge Conference*, Vancouver, British Columbia, Canada, 2017.
- [9] M. Sweeney, H. Rangwala, J. Lester, and A. Johri, "Next-term student performance prediction: a recommender systems approach," *Journal of Educational Data Mining*, vol. 8, no. 1, pp. 22–51, 2016.
- [10] G. Li, "Big data related technologies, challenges and future prospects," *Information Technology & Tourism*, vol. 15, no. 3, pp. 283–285, 2015.
- [11] P. Anand, "Big data is a big deal," *Journal of Petroleum Technology*, vol. 65, no. 4, pp. 18–21, 2015.
- [12] C. H. Miller, M. D. Sacchet, and I. H. Gotlib, "Support vector machines and affective science," *Emotion Review*, vol. 12, no. 4, pp. 297–308, 2020.
- [13] S. Kim and C. Kim, "Influence diagnostics in support vector machines," *Journal of the Korean Statistical Society*, vol. 49, no. 3, pp. 757–778, 2020.
- [14] J. A. Cook and S. Siddiqui, "Random forests and selected samples," *Bulletin of Economic Research*, vol. 72, no. 3, pp. 272–287, 2020.
- [15] S. Gil-Begue, C. Bielza, and P. Larrañaga, "Multi-dimensional Bayesian network classifiers: a survey," *Artificial Intelligence Review*, vol. 54, no. 1, pp. 519–559, 2021.
- [16] J. Barr, M. Littman, and M. desJardins, "Decision trees," *ACM Inroads*, vol. 10, no. 3, p. 56, 2019.
- [17] Y. P. Zhang, S. H. Wang, K. J. Xia, Y. Z. Jiang, and Q. J. Qian, "Alzheimer's disease multiclass diagnosis via multimodal neuroimaging embedding feature selection and fusion," *Information Fusion*, vol. 66, pp. 170–183, 2021.
- [18] Y. Jiang, X. Gu, D. Wu et al., "A novel negative-transfer-resistant fuzzy clustering model with a shared cross-domain transfer latent space and its application to brain CT image segmentation," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 18, no. 1, pp. 1–52, 2020.
- [19] Y. Jiang, Y. Zhang, C. Lin, D. Wu, and C.-T. Lin, "EEG-based driver drowsiness estimation using an online multi-view and transfer TSK fuzzy system," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 3, pp. 1752–1764, 2021.
- [20] Y. Zhang, Y. Jiang, L. Qi, M. Z. A. Bhuiyan, and P. Qian, "Epilepsy diagnosis using multi-view & multi-medoid entropy-based clustering with privacy protection," *ACM Transactions on Internet Technology*, vol. 21, no. 2, pp. 1–21, 2021.
- [21] X. Wang, A. Bao, Y. Cheng, and Q. Yu, "Multipath ensemble convolutional neural network," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 5, no. 2, pp. 298–306, 2021.
- [22] G. Dzhelyan and H. Cecotti, "Symmetrical filters in convolutional neural networks," *International Journal of Machine Learning and Cybernetics*, vol. 12, no. 7, pp. 2027–2039, 2021.
- [23] M. E. Valle and R. A. Lobo, "Hypercomplex-valued recurrent correlation neural networks," *Neurocomputing*, vol. 432, pp. 111–123, 2021.
- [24] M. Alameh, Y. Abbass, A. Ibrahim, G. Moser, and M. Valle, "Touch modality classification using recurrent neural networks," *IEEE Sensors Journal*, vol. 21, no. 8, pp. 9983–9993, 2021.
- [25] J. Macqueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, pp. 281–297, Berkeley, California, USA, 1967.

- [26] Z. Zhang, Q. Feng, J. Huang, Y. Guo, J. Xu, and J. Wang, "A local search algorithm for k-means with outliers," *Neurocomputing*, vol. 450, pp. 230–241, 2021.
- [27] K. Sando and H. Hino, "Modal principal component analysis," *Neural Computation*, vol. 32, no. 10, pp. 1–35, 2020.
- [28] A. Charpentier, S. Mussard, and T. Ouraga, "Principal component analysis: a generalized Gini approach," *European Journal of Operational Research*, vol. 194, no. 1, pp. 236–249, 2021.



## Research Article

# A Heterogeneous Ensemble Learning Model Based on Data Distribution for Credit Card Fraud Detection

Yalong Xie , Aiping Li , Liqun Gao , and Ziniu Liu 

*College of Computer, National University of Defense Technology, Changsha, Hunan 410073, China*

Correspondence should be addressed to Aiping Li; [liaiping@nudt.edu.cn](mailto:liaiping@nudt.edu.cn)

Received 11 June 2021; Revised 4 July 2021; Accepted 12 July 2021; Published 22 July 2021

Academic Editor: Shan Zhong

Copyright © 2021 Yalong Xie et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Credit card fraud detection (CCFD) is important for protecting the cardholder's property and the reputation of banks. Class imbalance in credit card transaction data is a primary factor affecting the classification performance of current detection models. However, prior approaches are aimed at improving the prediction accuracy of the minority class samples (fraudulent transactions), but this usually leads to a significant drop in the model's predictive performance for the majority class samples (legal transactions), which greatly increases the investigation cost for banks. In this paper, we propose a heterogeneous ensemble learning model based on data distribution (HELMDD) to deal with imbalanced data in CCFD. We validate the effectiveness of HELMDD on two real credit card datasets. The experimental results demonstrate that compared with current state-of-the-art models, HELMDD has the best comprehensive performance. HELMDD not only achieves good recall rates for both the minority class and the majority class but also increases the savings rate for banks to 0.8623 and 0.6696, respectively.

## 1. Introduction

With the rapid development of mobile internet and e-commerce technologies, online payment tools such as credit cards are welcomed by more and more people. While credit cards bring convenience to customers, they also expose cardholders and banks to potential fraud risks [1, 2]. Credit card fraud is a global problem. The Nilson report found that by 2023, the worldwide fraud loss is expected to reach \$35.67 billion annually [3]. Fraud prevention and fraud detection are two main ways to combat credit card fraud [4]. Fraud prevention consists of a series of rules, procedures, and protocols. Commonly used technologies in fraud prevention include secure payment gateways, intrusion detection systems, and firewalls [5]. Fraud detection takes place after the fraud prevention mechanism has been breached [4], which means that fraud detection is the last line of defense to ensure the security of credit card transactions. Banks have to invest considerable money to optimize their fraud detection system [6], due to the need to protect cardholder's funds and their own business reputation.

Data mining and machine learning are widely used technologies in financial fraud detection [7–9]. As early as 1998, researchers had begun to build CCFD systems based on machine learning techniques [10]. After more than two decades of development, researchers have proposed many different methods and models [2, 11]. In machine learning terms, CCFD is a typical binary classification problem. The detection system is aimed at determining whether the current transaction is either legal (the transaction was made by the cardholder) or fraudulent (the transaction was made by an unauthorized person) based on historical transaction data [12]. Various methods have been proposed to tackle this problem, including supervised learning, unsupervised learning, and semisupervised learning. In supervised learning, the historical transaction data (training data) are labeled with known outcomes. Commonly used supervised learning models include Hidden Markov Model (HMM) [13], Logistic Regression (LR) [14], Support Vector Machine (SVM) [15],  $K$ -nearest neighbors (KNN) [16], Bayesian Networks (BN) [17], Decision Tree (DT) [18], random forest (RF) [19], and Artificial Neural Network (ANN) [20]. Conversely, the

historical transaction data used in unsupervised learning models (ULMs) are unlabeled. ULMs judge whether transactions are fraudulent by observing the distribution of current and historical transaction data. Commonly used ULMs include artificial immune systems [21] and self-organizing maps [22]. Semisupervised learning models are a combination of supervised and unsupervised learning models, which use some labeled data in combination with a large amount of unlabeled data. This can help banks reduce the cost of labeling large volumes of transaction data [12, 23].

In the real world, the proportion of fraudulent transactions (minority class) is much lower than that of legitimate transactions (majority class), which means that the distribution of credit card transaction data is highly imbalanced, and this increases the difficulty of fraud detection [15, 24]. Most standard classifiers have poor performance on imbalanced data, especially for the minority class [25]. Resampling is a widely used method to address the problem of imbalanced classification data. Several resampling algorithms have been proposed to improve the recognition performance of classifiers for the minority class [26–28]. However, the disadvantage of the resampling method is that it significantly reduces the performance of classifiers for the majority class. For CCFD, this means that a large number of legal transactions are misclassified as fraudulent, which will significantly increase the investigation costs. Therefore, it is critical to build a CCFD model with strong recognition performance in both the minority and majority classes.

To address the above issues, we propose a new kind of heterogeneous ensemble learning model based on data distribution (HELMDD) for credit card fraud detection. The core idea is to incorporate a resampling method based on the distribution of data (RMDD). To reduce information loss in the majority class and improve the performance of the base classifiers, RMDD applies KNN and  $K$ -Means algorithms to obtain samples from the majority class, which retain its diversity and boundary contours. Finally, balanced subsets for training the base classifiers are obtained by pairing majority and minority class training subsets.

The main contributions of our study are as follows:

- (1) We design a new undersampling method based on the distribution of majority class samples, RMDD, which can reduce information loss within the majority class
- (2) We design a novel combination based on heterogeneous ensemble learning and our RMDD resampling method to obtain better prediction performance in highly imbalanced credit card transaction datasets
- (3) Experimental results on two real credit card fraud datasets demonstrate that the proposed model can achieve better performance

## 2. Literature Review

*2.1. Credit Card Fraud Detection Model.* Credit card datasets contain detailed information about each transaction, such as account number, transaction amount, time, location, and

merchant category. We can construct a model to determine whether a transaction is fraudulent or not by expressing the transaction-related information as vectors and calculating their similarity. Singh and Jain [29] reviewed literature on CCFD and summarized the topical issues in current research, such as datasets, evaluation matrices, and the advantages and disadvantages of different models. Armel and Zaidouni [30] compared and analyzed the effectiveness of simple anomaly detection using DT, RF, and Naive Bayes (NB) in CCFD through a series of experiments. Sohony et al. [4] found that RF enables higher accuracy in predicting legal transaction instances and a Feedforward Neural Network (FNN) achieves higher accuracy in predicting instances of fraudulent transactions. Consequently, they proposed an ensemble learning model based on RF and FNN.

Deep learning for CCFD has been discussed in several works [20, 31, 32]. Rushin et al. [20] conducted comparative experiments on deep learning, LR, and Gradient Boosted Tree (GBT) with a dataset containing approximately 80 million account level transactions. The results showed that the performance of deep learning models is better than the GBT and LR. Kim et al. [31] proposed a champion-challenger framework that includes deep learning and ensemble learning and evaluated it on a large transaction dataset taken from a major card issuing company in South Korea. Li et al. [32] proposed a deep representation learning model based on a full center loss function, which considers both distances and angles among different features.

Some studies have made improvements in feature engineering methods for credit card transaction data. Zhang et al. [24] proposed a feature engineering method based on homogeneity-oriented behavior analysis and then used a deep belief network for learning the extracted features. Lucas et al. [33] proposed an HMM-based feature engineering strategy that could incorporate sequential knowledge in the transactions in the form of HMM-based features, which enabled a nonsequential RF classifier to make use of the sequential information. Wu et al. [34] proposed a new feature engineering method to detect fraudulent cash-out of credit cards that considers both snapshot and dynamic behavioral patterns of cardholders and conducted a comparative experiment with the feature extraction method based on Whitrow's strategy. Vlasselaer et al. [35] proposed a feature engineering method based on the network structure of cardholders and merchants and then calculated a time-dependent suspiciousness score for each network object.

Many other approaches have been used recently in the identification of credit card fraud. Gianini et al. [36] proposed a method of rule pool management based on game theory in which the system distributes suspicious transactions for manual investigation while avoiding the need to isolate the individual rules. Based on generative adversarial networks, Fiore et al. [37] proposed a method to generate simulated fraudulent transaction samples to improve the effectiveness of classification models. Carcillo et al. [38] proposed a scalable real-time CCFD framework that could deal with imbalance and feedback latency based on big data tools such as Spark. Their work provides a reference for real-time detection in massive credit card transaction data.

**2.2. Imbalanced Data Learning Methods.** Imbalanced distribution of data (class imbalance) has a great impact on the performance of classification models, reducing the accuracy of prediction in the minority class [25]. Some effective solutions for class imbalanced data have been proposed by many researchers. These solutions can be arranged into two groups: data level and algorithm level [2].

Resampling is a simple and efficient way to address the problem of class imbalance at the data level. Current resampling strategies can be divided into those that oversample the minority class samples and those that undersample the majority class samples. Commonly used oversampling methods include Random Oversampling (ROS), Synthetic Minority Oversampling Technique (SMOTE) [39], and Borderline-SMOTE [40]. For a highly imbalanced credit card transaction dataset, oversampling generates many minority class samples (fraudulent transactions). Although this can increase the learning weight of the classification model for minority class samples, it also increases computational complexity and generates many noise samples, which will reduce the predictive performance for the majority class (legal transactions). Commonly used undersampling methods include Random Undersampling (RUS), one-sided dynamic undersampling [41], and neighborhood-based undersampling [42]. The undersampling approach involves deleting a large number of majority class samples. This improves the computational efficiency of the classification model but may result in the loss of important information from the majority class samples, which can increase the false-positive rate of the classification model and lead to additional investigation costs for the banks.

Cost-sensitive learning technology is often used to address the problem of imbalanced datasets at the algorithm level. These learning models introduce some constraints and weights through a cost matrix based on the loss function of conventional learning models, which causes models to shift to a smaller total cost. The advantage of cost-sensitive learning technology is that it does not generate or add new information, thereby avoiding the introduction of external noise into the classification model. The disadvantage of cost-sensitive learning technology is that the establishment of the cost matrix needs to be estimated by business experts and cannot be calculated accurately. Commonly used cost-sensitive learning models include cost-sensitive SVM [43], cost-sensitive LR [44], and cost-sensitive DT [18].

Akila and Reddy [45] proposed a cost-sensitive risk-induced Bayesian inference bagging model for CCFD to help card issuers reduce costs. They verified the effectiveness of this model on a dataset from a Brazilian bank. Nami and Shajari [46] proposed a two-stage detection algorithm to address class imbalance in payment card fraud detection. The first stage extracts the relevant features from the transaction data and the second stage extracts the recent transaction behavioral characteristics of cardholders. In the second stage, a cost-sensitive dynamic random forest model is used to improve classification performance.

### 3. Methodology

In this section, we introduce the proposed heterogeneous ensemble learning model based on data distribution (HELMDD) in details, which consists of two main components. The first is a resampling method based on data distribution (RMDD), as illustrated in Figure 1. RMDD undersamples the majority class based on the data distribution of the majority samples and creates several balanced training subsets by using KNN and *K*-Means. The second one is a framework based on a heterogeneous ensemble learning model (HELM), as illustrated in Figure 2. HELM is a framework that integrates seven kinds of heterogeneous classification models (LR, SVM, NB, DT, RF, AdaBoost, and XGBoost) in the bagging method.

**3.1. KNN.** KNN is a widely used unsupervised learning method. KNN can predict the category of samples by calculating the Euclidean distance between different points. The formula for calculating the Euclidean distance between points  $(x_i, y_j)$  is shown in the following equation:

$$d_{ij} = \sqrt{\sum_{k=1}^n (x_{ik} - y_{jk})^2} \quad (k = 1, 2, \dots, n). \quad (1)$$

RMDD divides the majority class samples of the training dataset into a subset of boundary samples and a subset of ordinary samples using a KNN algorithm. A selection of samples is then drawn from each of these subsets to create several new balanced datasets that contain cases from both the majority and minority classes. The advantage of this method is that the new balanced training subsets retain some of the boundary features of the majority class from the original training dataset, which can reduce information loss in these critical boundary cases.

**3.2. *K*-Means.** *K*-Means is a popular unsupervised clustering algorithm. Taking dataset  $S$  and the number of classes  $k$  as inputs, the *K*-Means algorithm is aimed at dividing  $S$  into  $k$  subsets quickly. Specifically, *K*-Means randomly select  $k$  samples as initial clusters. Then, for each sample  $s$  in the dataset, the Euclidean distance  $d_{sj}$  between sample  $s$  and the centroid of  $k$  different clusters is calculated. If the distance between  $s$  and the centroid of cluster  $i$  is the shortest,  $s$  is assigned to cluster  $i$ . The third step is to calculate the average value of samples in cluster  $i$  and update the centroid of cluster  $i$ . The second and third steps are repeated until the difference between the old centroid and the new centroid is less than a preset threshold. After the algorithm is executed, we can obtain the data distribution of the majority class samples.

In the task of imbalanced classification, undersampling methods can increase the learning weight of minority class samples, which helps classification models to attain a higher recall rate. RUS is the most widely used method, but a significant defect of RUS is that it discards a large number of samples from the majority class, which may increase the false-positive rate of classification models. During the undersampling process, if we do not consider the distribution of

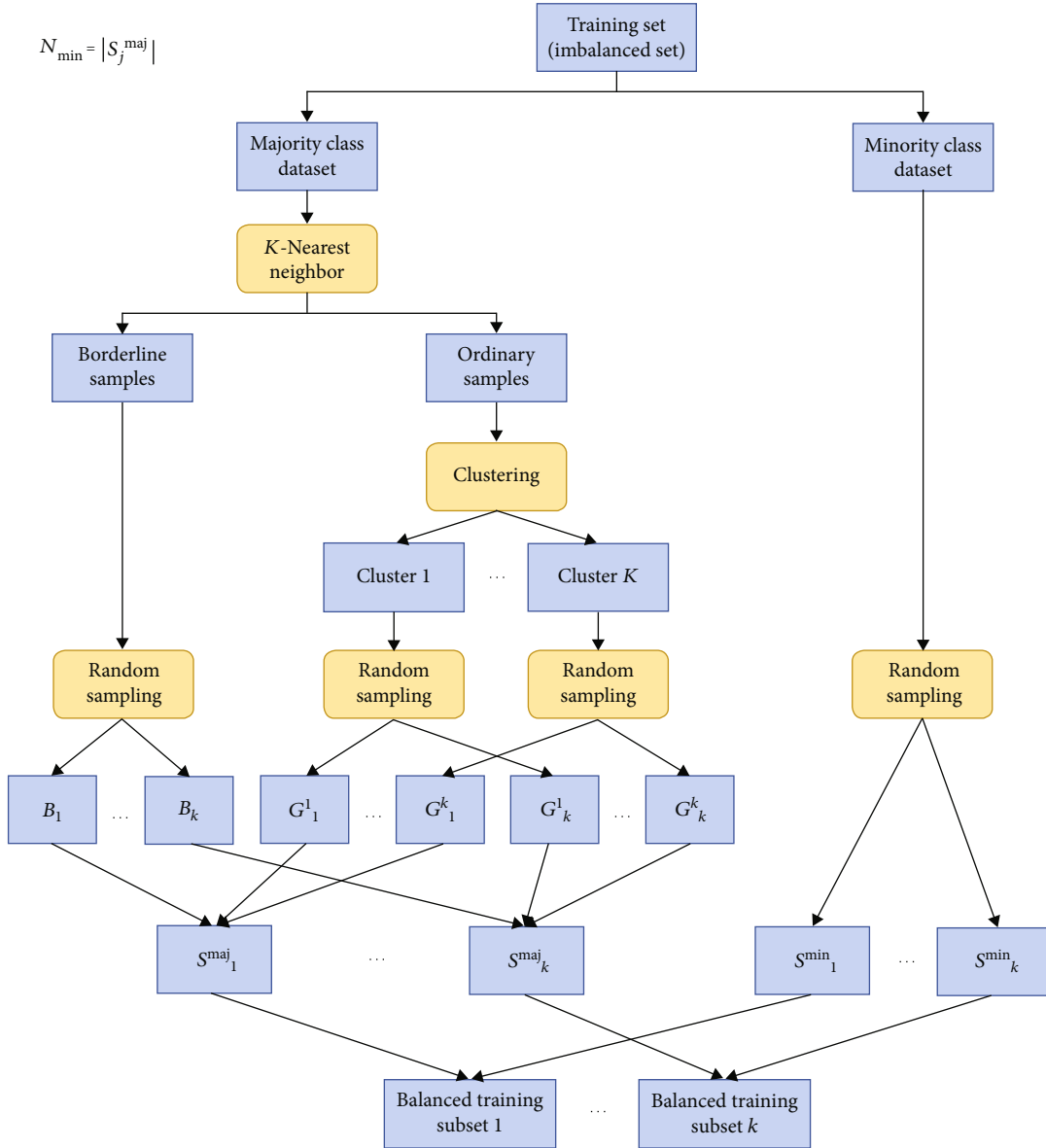


FIGURE 1: Flowchart of RMDD resampling algorithm, including three classic algorithms: KNN, K-Means, and RUS.

majority class samples, the selected samples cannot represent main features of the majority class and thus could decrease the performance of the base classifiers. In our method, we divide the majority class into  $k$  clusters using the  $K$ -Means algorithm and then randomly sample from each cluster in different proportions. This resampling method fully considers the distribution of majority class samples and thus can better retain the main features of these cases.

**3.3. RMDD Resampling Method.** RMDD is an undersampling algorithm that fully considers data distribution, which has three components. The first is to sample the minority class. Due to the highly imbalanced distribution in a CCFD dataset, we use all the minority class samples to improve recognition ability for the minority class of the base classifier. In the second part, we undersample the majority class to generate mul-

iple subsets so that the number of majority class samples is the same as the number of minority class samples, which forms the core of the RMDD algorithm. The third part is to generate several balanced subsets to provide training data for the base classifiers by merging the minority class samples and the subset of majority class samples. The flowchart of the RMDD resampling algorithm is shown in Figure 1.

The second part above consists of the following 4 steps:

- (1) We divide the majority class samples into a boundary sample set  $B$  and an ordinary sample set  $G$  by the KNN algorithm. Then, for any sample  $x$  in the majority class (labeled 0), we find  $K$ -nearest neighbor samples to  $x$  from the training set. If there are more than  $k/2$  neighboring samples with a label of 1, then  $B = B \cup \{x\}$ , otherwise  $G = G \cup \{x\}$

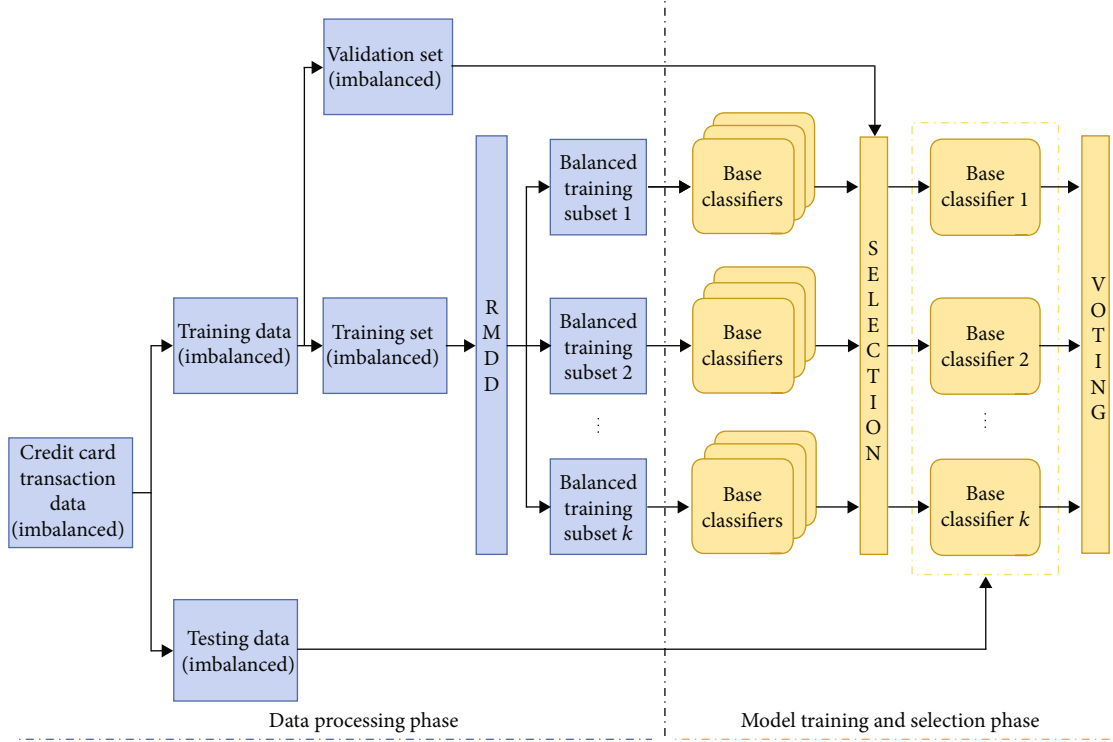


FIGURE 2: Framework of HELM. Preprocessing original data by RMDD resampling algorithm and selecting best classifiers from different base classifiers by AUC score.

- (2) Random sampling with replacement is used to divide sample  $B$  into  $k$  subsets.  $N_b$  represents the number of samples in set  $B$ . The formula for calculating the number of samples in each subset is shown as

$$N_{bi} = \lambda_i N_b, \quad i \in [1, 2, \dots, k], \lambda_i \in [0, 1], \quad (2)$$

$$\sum_{i=1}^k \lambda_i = 1. \quad (3)$$

$\lambda_i$  is a weight parameter, which is used to adjust the number of boundary samples in each subset

- (3) Using the  $K$ -Means algorithm, we divide ordinary sample set  $G$  into  $k$  clusters and sample with replacement from each cluster with different sampling rate. The sampling rate in the  $i$ th cluster is calculated from Equation (4), where  $N_i$  represents the number of samples in the  $i$ th cluster, and  $N_g$  represents the number of ordinary samples in set  $G$ .  $N_{\min}$  represents the number of minority samples, and  $N_{gi}$  represents the number of samples randomly selected from the  $i$ th cluster and can be calculated from Equation (5).

$$SR_i = \frac{N_i}{N_g}, \quad i \in [1, 2, \dots, k], \quad (4)$$

$$N_{gi} = (N_{\min} - N_{bi})SR_i \quad (5)$$

- (4) Combining boundary sample subset  $B_j$  with ordinary sample subset  $G_j$  to construct a subset  $S_j^{\text{maj}}$  of majority class samples. The corresponding calculation formula is as follows:

$$S_j^{\text{maj}} = \{B_j, G_j^1, G_j^2, \dots, G_j^k\}, \quad j \in [1, 2, \dots, k]. \quad (6)$$

The number of samples in the majority class is equal to the number in the minority class, that is,  $N_{\min} = |S_j^{\text{maj}}|$

**3.4. Framework of HELM.** To improve classification performance, we propose a heterogeneous ensemble learning model (HELM) framework, as shown in Figure 2. The HELM framework uses data resampling and ensemble learning technologies to address the problem of imbalanced data in CCFD. Through the training and screening of multiple heterogeneous base classifiers, we improve the robustness of HELM as well as avoid reliance on a single classifier. The HELM framework consists of two phases: (1) the data processing phase and (2) the model training and selection phase.

**3.4.1. Data Processing Phase.** The main task of this phase is to preprocess the original credit card transaction dataset, including feature selection, data normalization, dataset division, and resampling. First, we divide the original dataset into

a training set and a test set. The training set is used for estimating parameters of the classification model, and the test set is used for evaluating the trained classification model. We further divide the training set into a training subset and a validation subset and use the RMDD algorithm to resample the training subset, which divides the highly imbalanced training subset into  $k$  balanced subsets for training base classifiers. The RMDD algorithm fully considers the distribution of majority class samples. Boundary samples are distinguished from ordinary samples by applying the KNN algorithm. Ordinary samples are grouped into  $k$  classes by the  $K$ -Means algorithm. Through these two algorithms, we can build several balanced subsets and ensure that each balanced training subset contains a certain ratio of boundary samples and ordinary samples from each cluster. The advantage of this is that more feature information of samples in the majority class can be preserved while generating new balanced training subsets. In addition, the introduction of boundary samples retains some of boundary contours from the original dataset in the new balanced subset, which can help to improve classification performance.

**3.4.2. Model Training and Selection Phase.** When the preprocessing phase is completed, we have  $k$  balanced training subsets. For each subset, we use seven different base classifiers for training, including LR, SVM, NB, DT, RF, AdaBoost, and XGBoost. Then, we use an imbalanced validation subset to obtain Area Under the Curve (AUC) score for each base classifier and select the base classifier with the best AUC score as recommended classifier for that subset. Finally, we obtain an ensemble learning model with heterogeneous or isomorphic recommended classifiers that are trained with other subsets. For samples in the test dataset, each recommended classifier will give an initial prediction; then, the final prediction result is generated through a voting method across each recommended classifier. In the model selection phase, we use AUC score as the selection condition because AUC score takes into account both the prediction accuracies of the majority and minority classes at the same time, which gives us a good compromise between the accuracy and recall metrics for the classification model. In credit card fraud prediction, misclassification of legitimate transactions as fraudulent transactions or mis classification of fraudulent transactions as legitimate transactions will incur costs for banks and customers, such as loss of transaction amount, manual investigation costs, etc. Therefore, by comparing the AUC score of multiple base classifiers and selecting the base classifier with the best AUC score to build an ensemble model, we can effectively improve prediction performance and reduce economic losses for cardholders and banks.

The HELM framework can be deployed in a distributed manner. Base classifier training tasks on different subsets in HELM can be assigned to different cluster nodes. Each node can perform model training in parallel during periods of low credit card transactions (such as the early morning). Since the proportion of fraudulent transactions is very low, the balanced training subset space generated by the RMDD algorithm is quite small, which can significantly reduce the training time of base classifiers. Compared with other tradi-

tional methods, our HELM framework can significantly reduce model training and deployment time by reducing the training sample space and facilitating the application of parallel computing technologies.

## 4. Experiments

**4.1. Dataset Description.** In this paper, we use two real credit card transaction datasets: one from Kaggle (public dataset) and one from a bank in China (our private dataset). The detailed statistics are shown in Table 1.

- (1) Kaggle dataset [47]. This dataset is composed of credit card transaction records of European cardholders in September 2013. The time span of these transactions is two days, and each transaction record contains 30 features. Due to privacy considerations, 28 features were encoded by Principal Component Analysis (PCA), except for two features: transaction time and amount. This dataset contains a total of 284,807 instances, of which 492 are minority class samples (fraudulent transactions). The fraud rate of this dataset is 0.173%, which indicates that the dataset is highly imbalanced
- (2) Our private dataset. This dataset is provided by a bank in China and contains credit card transaction records of customers on a typical day in May 2017. Each instance has 23 features, including some personal information of cardholders (such as age, gender, marital status, and education level) and transaction-related features (such as transaction amount, time, and merchant number). This dataset contains 24,024 instances, including 660 fraud instances. The fraud rate of this dataset is 2.747%.

**4.2. Performance Measures.** Confusion matrix provides helpful information regarding the actual labels and predicted labels proposed by the classification model. The confusion matrix used in this study is shown in Table 2. Due to the highly imbalanced phenomenon of our credit card datasets, widely used evaluation indexes (such as accuracy and precision) do not fully represent the performance of classification models. For example, if we classify all samples in the Kaggle dataset as legitimate transactions, the accuracy will be close to 98%; it is clear that this prediction model is not a good classification model. Therefore, we choose *Fra\_Recall* (fraud class recall), *Leg\_Recall* (legal class recall), *G-mean*, *AUC*, and *savings rate* [48] to evaluate the model.

*Fra\_Recall* and *Leg\_Recall* are calculated by Equations (7) and (8), respectively. The larger the *Fra\_Recall* value, the higher the proportion of fraudulent transactions that are identified by the classification model, and the more fraud losses that can be avoided for banks and cardholders. The larger the *Leg\_Recall* value, the higher the proportion of legitimate transactions that is identified by the classification model, and the greater the investigation costs that can be saved for banks. The ideal model is that *Fra\_Recall* and *Leg\_Recall* are close to 1 at the same time. *G-mean* and *AUC* are very important measures that are widely used in

TABLE 1: Dataset description.

Dataset	Instances	Features	Fraud instances (minority class)	Legal instances (majority class)	Fraud ratio
Kaggle dataset	284807	30	492	284315	0.173%
Our private dataset	24024	23	660	23364	2.747%

TABLE 2: Confusion matrix.

	Actual fraud (positive)	Actual legal (negative)
Predicted fraud (positive)	True positive (TP)	False positive (FP)
Predicted legal (negative)	False negative (FN)	True negative (TN)

model evaluation studies in the presence of imbalanced data. The larger the  $G$ -mean and AUC value, the better the performance of the classification model.  $G$ -mean and AUC can be calculated by Equations (9) and (10).

$$\text{Fra\_Recall} = \text{sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (7)$$

$$\text{Leg\_Recall} = \text{specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}, \quad (8)$$

$$G\text{-mean} = \sqrt{\text{sensitivity} \times \text{specificity}}, \quad (9)$$

$$\text{AUC} = 1 - \frac{1}{|D^+| \cdot |D^-|} \sum_{x^+ \in D^+} \sum_{x^- \in D^-} \left( \mathbf{I}(f(x^+) < f(x^-)) + \frac{1}{2} \mathbf{I}(f(x^+) = f(x^-)) \right). \quad (10)$$

$D^+$  and  $D^-$  denote the collection of fraudulent transactions and legitimate transactions, respectively.

The savings rate is an indicator that banks attach great importance to, because it is always used to quantify the economic benefits that fraud detection models can create for banks. The CCFD cost matrix [48] is shown in Table 3. Among them,  $y_i$  is the actual label for transaction  $i$ , and  $y'_i$  is the predicted label for transaction  $i$  given by classifier  $f$ . If a transaction is predicted to be a fraudulent transaction (TP or FP), the bank needs to investigate the transaction incurring a cost of  $C_a$ . Conversely, if the transaction is predicted to be legitimate (TN or FN), there is no investigation cost, but in the case of FN, the loss of the bank is equal to the transaction amount  $\text{Amt}_i$ . If no classifier is used for CCFD, the total loss of the bank is calculated by Equation (11). The proportion of cost saved for the bank by using classifier  $f$  is calculated by Equation (12).

$$C_{\text{total}} = \sum_{i=1}^n y_i \text{Amt}_i, \quad (11)$$

$$\text{Savings}(f) = \frac{\left( \sum_{i=1}^n y_i y'_i \text{Amt}_i - y_i' C_a \right)}{C_{\text{total}}} \quad (12)$$

**4.3. Experimental Design.** To evaluate the effectiveness of the HELMDD model, we conducted experiments on two real

credit card datasets and compared the proposed model with several competing approaches. Most of models can be divided into two categories: independent model and ensemble learning model. The independent model we used in our experiment includes LR, SVM, NB, and DT. The ensemble learning model we used in our experiment includes RF, AdaBoost, and XGBoost. In addition, we also combined these models with different resampling methods, such as SMOTE and RUS.

## 5. Experimental Results and Discussion

To directly compare with previous works, we evaluate our model using 10-fold cross-validation similar to prior approaches on the two datasets. The experimental results for each classification model on the Kaggle dataset and our private dataset are shown in Tables 4 and 5, respectively. For convenience of comparison, we have also presented the data of Tables 4 and 5 in histogram form, as shown in Figures 3 and 4. In Tables 4 and 5, numbers in italic indicate the best values of the model in the corresponding evaluation measure.

For the Kaggle dataset, we compare the proposed model with several competing approaches and show the results in Table 4. From the results, we can observe the following:

- (1) In the case of the same classification model, those implementing resampling methods to preprocess the training subset achieved better performance than models with the original imbalanced training subset. Fra\_Recall, AUC, and G-mean have different degrees of improvement. For example, Fra\_Recall increased from 0.0235 (DT model with SMOTE method) to 0.2353 (LR model with RUS method), AUC increased from 0.0023 (AdaBoost model with SMOTE method) to 0.0275 (NB model with RUS method), and G-mean increased from 0.0018 (DT model with RUS method) to 0.1239 (LR model with RUS). The main reason is that preprocessing the original imbalanced dataset with SMOTE or RUS method helps models to improve the learning rate of fraudulent transaction instances and therefore enhances the ability to identify fraudulent transactions

TABLE 3: Credit card fraud cost matrix.

	Actual fraud (positive) $y_i = 1$	Actual legal (negative) $y_i = 0$
Predicted fraud (positive) $y'_i = 1$	$C_{TP_i} = C_a$	$C_{FP_i} = C_a$
Predicted legal (negative) $y'_i = 0$	$C_{FN_i} = Amt_i$	$C_{TN_i} = 0$

TABLE 4: Results on the Kaggle dataset.

Model	Rebalancing technique	Fra_Recall	Leg_Recall	AUC	G-mean	Savings
LR	Imbalanced	0.6176	0.9942	0.9500	0.7836	0.5980
	SMOTE	0.8412	0.9755	0.9571	0.9058	0.7591
	RUS	0.8529	0.9657	0.9630	0.9075	0.7399
SVM	Imbalanced	0.8059	0.9963	0.9559	0.8961	0.7893
	SMOTE	0.8824	0.9700	0.9678	0.9252	0.7829
	RUS	0.8529	0.9605	0.9697	0.9051	0.7237
NB	Imbalanced	0.7001	0.9996	0.9413	0.8365	0.6943
	SMOTE	0.8012	0.9920	0.9539	0.8908	0.7700
	RUS	0.8059	0.9783	0.9688	0.8879	0.7329
DT	Imbalanced	0.7647	0.9996	0.8821	0.8743	0.7585
	SMOTE	0.7882	0.9965	0.8924	0.8863	0.7723
	RUS	0.8471	0.9062	0.9061	0.8761	0.5477
RF	Imbalanced	0.7588	0.9999	0.9614	0.8711	0.7538
	SMOTE	0.8235	0.9998	0.9709	0.9074	0.8178
	RUS	0.8588	0.9703	0.9719	0.9129	0.7604
AdaBoost	Imbalanced	0.7765	0.9998	0.9601	0.8811	0.7708
	SMOTE	0.8706	0.9850	0.9624	0.9260	0.8182
	RUS	0.8824	0.9473	0.9694	0.9143	0.7117
XGBoost	Imbalanced	0.7706	0.9999	0.9649	0.8778	0.7656
	SMOTE	0.8765	0.9885	0.9700	0.9308	0.8194
	RUS	0.9001	0.9663	0.9725	0.9325	0.7886
HELMDD	RMDD	0.8882	0.9903	0.9853	0.9379	0.8623

(2) In the case of applying the same resampling method, AUC obtained by ensemble learning models are generally better than those from independent learning models. As shown in Table 4, the highest AUC obtained by independent learning models with three different resampling methods are 0.9559 (imbalanced data), 0.9678 (SMOTE method), and 0.9697 (RUS method), while the average AUC obtained by ensemble learning models with three different sampling methods are 0.9621 (imbalanced data), 0.9678 (SMOTE method), and 0.9713 (RUS method), so the ensemble learning models are slightly better than the independent learning models. This is because ensemble learning models are strengthened by using multiple weak classification models. Compared with

independent models, ensemble models can obtain a smaller deviation and better generalization ability

(3) For the same classification model, those using RUS to preprocess the training dataset achieve better Fra\_Recall and AUC than models based on SMOTE. For example, Fra\_Recall increases from 0.0047 (NB model) to 0.0589 (DT model), and AUC increases from 0.0010 (RF model) to 0.0149 (NB model). However, we cannot ignore that Leg\_Recall decreases by 0.0095 (SVM model) to 0.0903 (DT model). This is because the RUS method discards many legitimate transaction samples and leads to an improvement in the identification of fraudulent transactions while increasing the false prediction rate for legitimate transactions



TABLE 5: Results on our private dataset.

Model	Rebalancing technique	Fra_Recall	Leg_Recall	AUC	G-mean	Savings
LR	Imbalanced	0.5928	0.8145	0.7334	0.6949	0.5414
	SMOTE	0.5825	0.7431	0.7058	0.6579	0.5130
	RUS	0.6959	0.6483	0.7250	0.6717	0.6014
SVM	Imbalanced	0.5155	0.7679	0.6567	0.6291	0.4527
	SMOTE	0.6959	0.6068	0.6881	0.6498	0.5909
	RUS	0.5773	0.8751	0.7509	0.7108	0.5415
NB	Imbalanced	0.3505	0.9428	0.7076	0.5749	0.3335
	SMOTE	0.5722	0.8055	0.7245	0.6789	0.5186
	RUS	0.5722	0.8038	0.7274	0.6782	0.5182
DT	Imbalanced	0.0876	0.9659	0.5268	0.2909	0.0783
	SMOTE	0.1598	0.9072	0.5335	0.3807	0.1350
	RUS	0.6443	0.6017	0.6230	0.6226	0.5384
RF	Imbalanced	0.5773	0.7958	0.7252	0.6778	0.5213
	SMOTE	0.5412	0.7126	0.6564	0.6210	0.4642
	RUS	0.6186	0.7861	0.7528	0.6973	0.5598
AdaBoost	Imbalanced	0.6186	0.8047	0.7465	0.7055	0.5645
	SMOTE	0.6031	0.7360	0.7101	0.6662	0.5316
	RUS	0.6392	0.7531	0.7581	0.6938	0.5718
XGBoost	Imbalanced	0.6649	0.7387	0.7543	0.7008	0.5937
	SMOTE	0.6082	0.7398	0.7014	0.6708	0.5377
	RUS	0.6753	0.7693	0.7624	0.7208	0.6118
HELMDD	RMDD	0.7213	0.7985	0.7941	0.7589	0.6696

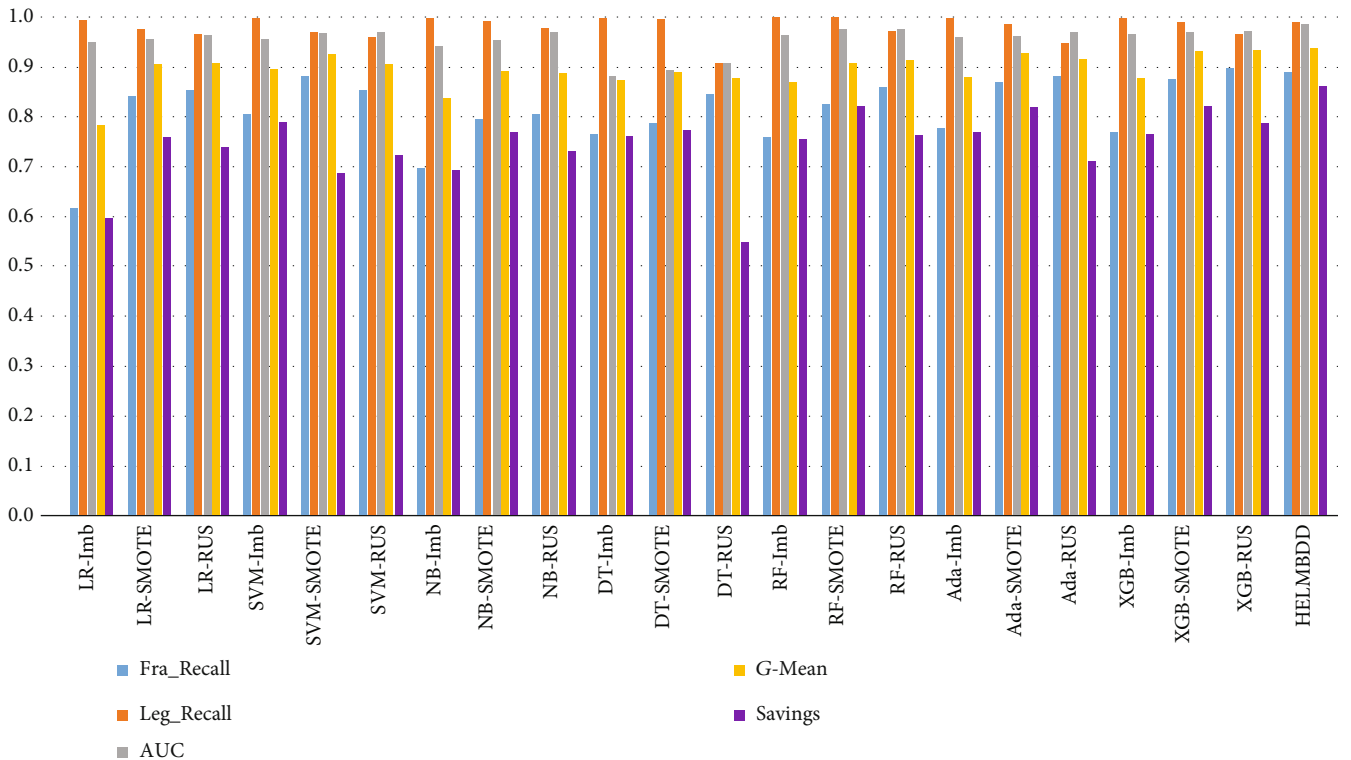


FIGURE 3: Histogram for the mean of model evaluation measures on the Kaggle dataset.

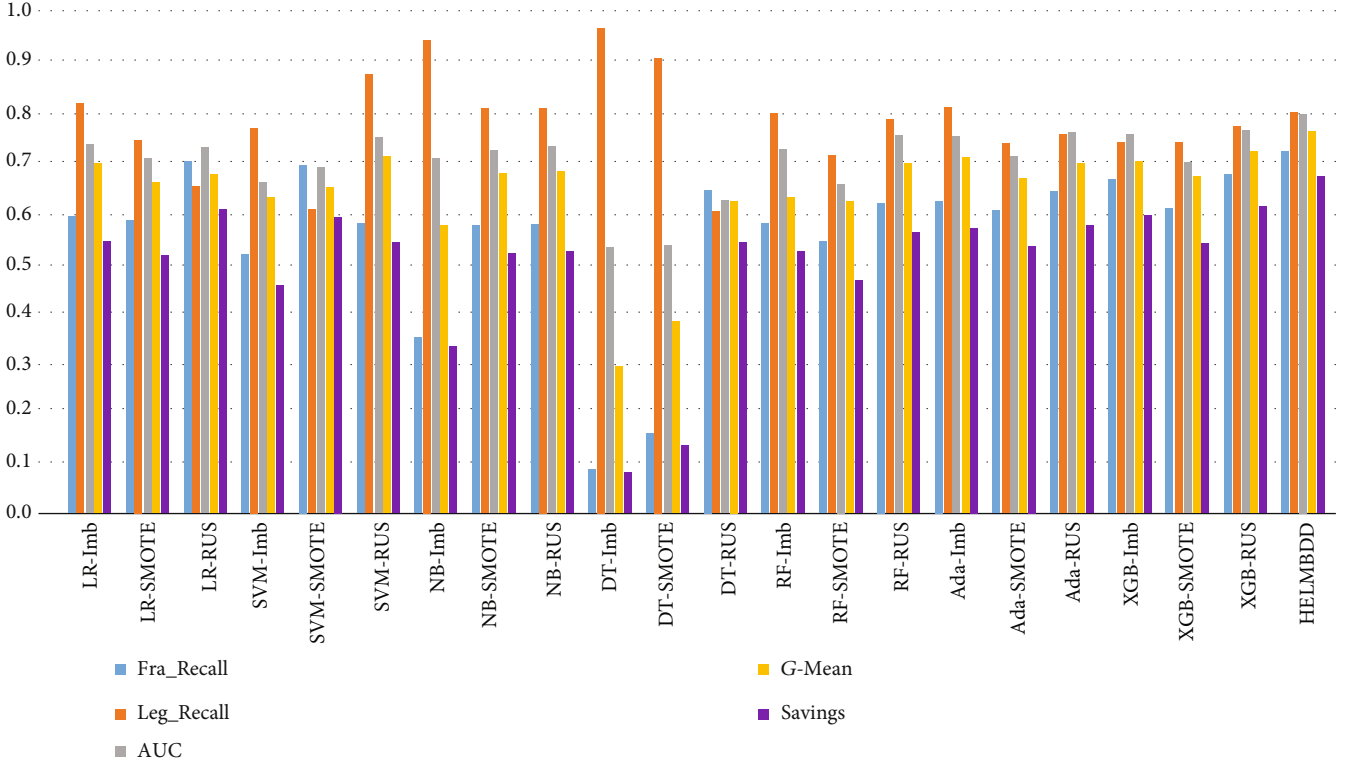


FIGURE 4: Histogram for the mean of model evaluation measures for our private dataset.

- (4) In terms of savings score, we have two findings: first, after resampling the original imbalanced training data with SMOTE method, the savings score of six classification models has been improved to varying degrees. For example, the XGBoost model has been increased from 0.7656 to 0.8194 and the LR model has been upgraded from 0.5980 to 0.7591. Second, when we use the RUS method to resample original data, the savings score of three classification models (SVM, DT, and AdaBoost) has been reduced by different degrees, such as the SVM model which has been reduced from 0.7893 to 0.7237, while the savings score of the other four classification models have been improved by different degrees, such as the XGBoost model which has been increased from 0.7656 to 0.7886. There may be two reasons for this: first, the savings score is highly correlated with the recognition rate of fraudulent transactions; SMOTE method can help classification models increase the recall rate of fraudulent transactions and reduce fraud losses of banks. Second, the RUS method discards many legitimate transaction samples, although it strengthens the learning of fraud samples and improves the recall rate of fraud transaction for the model, but it also leads to an increase in the false prediction rate of legitimate transactions and increases investigation cost for banks
- (5) Our HELMDD model proposed in this paper achieved the best AUC,  $G$ -mean, and savings scores,

which were 0.0128, 0.0054, and 0.0429 higher than previous state-of-the-art methods, respectively. The model showed good stability. While obtaining the second highest Fra\_Recall, it did not significantly reduce Leg\_Recall, thus ensuring that banks can achieve greater savings. The overall performance of HELMDD is better than the ensemble learning models (such as XGBoost, AdaBoost, and RF) with different resampling methods. This is because the RMDD resampling algorithm fully considers the distribution of legitimate transaction samples. Samples extracted from the boundary subset and multiple clusters fully retain the diversity and boundary contours of legitimate transaction samples. In addition, the selection mechanism of base classification models also helps to improve the overall performance of the framework

Table 5 presents the performance comparison between our approach and other competitive methods on our private dataset. From the results, we can observe the following:

- (1) Using the SMOTE method to resample the training dataset may not necessarily improve performance of the classification models and may even lead to a deterioration in classification performance. In Table 5, AUC and  $G$ -mean obtained by SVM, NB, and DT combined with SMOTE have been improved to varying degrees. For example, AUC increases from 0.0067 (DT model) to 0.0314 (SVM model), and  $G$ -mean

TABLE 6: Ablations on two datasets with different resampling and model ensemble methods.

Method	Kaggle dataset		Our private dataset	
	AUC	Savings	AUC	Savings
HELMDD	0.9853	0.8623	0.7941	0.6696
-RMDD	0.9764	0.8255	0.7737	0.6324
-HELM	0.9788	0.8472	0.7853	0.6582
Base	0.9725	0.7886	0.7624	0.6118

increases from 0.0207 (SVM model) to 0.1040 (NB model). However, the overall performance of LR, RF, AdaBoost, and XGBoost combined with SMOTE decreases to varying degrees, in which AUC decreases by 0.0276 (LR model) to 0.0688 (RF model), and  $G$ -mean decreases from 0.0300 (XGBoost model) to 0.0568 (RF model). This is possibly because through the SMOTE method, a large amount of minority sample noise is generated in the process of resampling for the training dataset, which decreases the performance of some classification models

- (2) Using RUS to resample the training dataset, apart from AUC of the LR model that dropped from 0.7334 to 0.7250, AUC obtained by the other six classification models manifests an improvement of 0.0081 (XGBoost model) to 0.0962 (DT model). This may be because the absence of new minority samples generated in the RUS process avoids the introduction of noise samples and improves the performance of the classification models
- (3) The HELMDD model proposed in this article achieved the best  $Fra\_Recall$ , AUC,  $G$ -mean, and savings scores, which were 0.0254, 0.0317, 0.0381, and 0.0578 higher than the corresponding measures of the previous state-of-the-art models, respectively. The validity and stability of HELMDD model were thus verified again

## 6. Ablation Study

We conduct an ablation study to investigate the effectiveness of our model components.

Table 6 shows the effects of the different resampling and model ensemble methods on AUC and savings scores. Here, -RMDD denotes using RUS instead of the RMDD resampling technique in HELMDD, and -HELM denotes using XGBoost instead of the seven heterogeneous models and ensemble in HELMDD. Base is the model generated by performing the above two ablations, which is the default XGBoost model with RUS resampling. For Kaggle and our private dataset, we observe that both RMDD and HELM are beneficial for identifying fraudulent transactions and controlling the cost of investigating fraudulent transactions. The reason is that two model components can significantly improve the recognition rate of fraudulent transactions without reducing the recognition rate of legitimate transactions.

## 7. Conclusions

In this paper, we propose a heterogeneous ensemble learning model based on data distribution (HELMDD) for the problem of the highly imbalanced data distribution encountered in CCFD. In our HELMDD model, we first propose an undersampling method, RMDD, based on the distribution of the majority class. RMDD divides the majority class into boundary samples and ordinary samples and then generates multiple balanced subsets based on the idea of clustering to train multiple base classifiers. The RMDD algorithm can maintain the classification boundary contours of the majority class and reduce the loss of sample information. Therefore, our model can obtain a higher majority class recall rate while also improving the minority class recall rate. In terms of model selection, we chose base classifiers that obtain the best AUC score in the balanced subset to generate an ensemble model, which helped to improve classification performance. Finally, we evaluate the proposed method on the Kaggle dataset and our private dataset. The results show that HELMDD achieves new state-of-the-art performance compared to other competing approaches.

## Data Availability

The labeled dataset used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare no conflicts of interest.

## Acknowledgments

This work was partially supported by the National Natural Science Foundation of China (Nos. 61732022, 61732004, 61672020, and 62072131) and the National Key R&D Program of China (Nos. 2017YFB0802204, 2019QY1406, and 2017YFB0803303).

## References

- [1] L. Zheng, G. Liu, C. Yan, and C. Jiang, "Transaction fraud detection based on total order relation and behavior diversity," *IEEE Transactions on Computational Social Systems*, vol. 5, no. 3, pp. 796–806, 2018.
- [2] S. Makki, Z. Assaghir, Y. Taher, R. Haque, M.-S. Hacid, and H. Zeineddine, "An experimental study with imbalanced classification approaches for credit card fraud detection," *IEEE Access*, vol. 7, pp. 93010–93022, 2019.
- [3] C. V. Priscilla and D. P. Prabha, "Influence of optimizing XGBoost to handle class imbalance in credit card fraud detection," in *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*, pp. 1309–1315, Tirunelveli, TN, India, August 2020.
- [4] I. Sohony, R. Pratap, and U. Nambiar, "Ensemble learning for credit card fraud detection," in *2018 Proceedings of the ACM India Joint International Conference on Data Science and Management of Data (CoDS-COMAD)*, pp. 289–294, Panaji, Goa, India, January 2018.

- [5] A. Abdallah, M. A. Maarof, and A. Zainal, "Fraud detection system: A survey," *Journal of Network and Computer Applications*, vol. 68, pp. 90–113, 2016.
- [6] J. Jurgovsky, M. Granitzer, K. Ziegler et al., "Sequence classification for credit-card fraud detection," *Expert Systems with Applications*, vol. 100, pp. 234–245, 2018.
- [7] N. Carneiro, G. Figueira, and M. Costa, "A data mining based system for credit-card fraud detection in e-tail," *Decision Support Systems*, vol. 95, pp. 91–101, 2017.
- [8] S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C. Westland, "Data mining for credit card fraud: a comparative study," *Decision Support Systems*, vol. 50, no. 3, pp. 602–613, 2011.
- [9] S. Carta, G. Fenu, D. R. Recupero, and R. Saia, "Fraud detection for E-commerce transactions by employing a prudential multiple consensus model," *Journal of Information Security and Applications*, vol. 46, pp. 13–22, 2019.
- [10] P. K. Chan and S. J. Stolfo, "Toward scalable learning with non-uniform class and cost distributions: a case study in credit card fraud detection," in *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 164–168, New York, NY, USA, August 1998.
- [11] S. J. Omar, K. Fred, and K. K. Swaib, "A state-of-the-art review of machine learning techniques for fraud detection research," in *Proceedings of the 2018 International Conference on Software Engineering in Africa (ICSE)*, pp. 11–19, Gothenburg, GOT, Sweden, May 2018.
- [12] F. Carcillo, Y.-A. L. Borgne, O. Caelen, Y. Kessaci, F. Oblé, and G. Bontempi, "Combining unsupervised and supervised learning in credit card fraud detection," *Information Sciences*, vol. 557, pp. 317–331, 2021.
- [13] V. Bhusari and S. Patil, "Study of hidden Markov model in credit card fraudulent detection," in *2016 World Conference on Futuristic Trends in Research and Innovation for Social Welfare (Startup Conclave)*, pp. 1–4, Coimbatore, India, February 2016.
- [14] S. Jha, M. Guillen, and J. C. Westland, "Employing transaction aggregation strategy to detect credit card fraud," *Expert Systems with Applications*, vol. 39, no. 16, pp. 12650–12657, 2012.
- [15] N. Rtayli and N. Enneya, "Enhanced credit card fraud detection based on SVM-recursive feature elimination and hyper-parameters optimization," *Journal of Information Security and Applications*, vol. 55, article 102596, 2020.
- [16] H. Wang, P. Zhu, X. Zou, and S. Qin, "An ensemble learning framework for credit card fraud detection based on training set partitioning and clustering," in *2018 IEEE SmartWorld, Ubiquitous Intelligence Computing, Advanced Trusted Computing, Scalable Computing Communications, Cloud Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UiC/ATC/CBD)*, pp. 94–98, Guangzhou, GZ, China, October 2018.
- [17] S. Panigrahi, A. Kundu, S. Sural, and A. K. Majumdar, "Credit card fraud detection: a fusion approach using Dempster-Shafer theory and Bayesian learning," *Information Fusion*, vol. 10, no. 4, pp. 354–363, 2009.
- [18] Y. Sahin, S. Bulkan, and E. Duman, "A cost-sensitive decision tree approach for fraud detection," *Expert Systems with Applications*, vol. 40, no. 15, pp. 5916–5923, 2013.
- [19] S. Bagga, A. Goyal, N. Gupta, and A. Goyal, "Credit card fraud detection using pipeling and ensemble learning," *Procedia Computer Science*, vol. 173, pp. 104–112, 2020.
- [20] G. Rushin, C. Stancil, M. Sun, S. Adams, and P. Beling, "Horse race analysis in credit card fraud—deep learning, logistic regression, and gradient boosted tree," in *2017 Systems and Information Engineering Design Symposium (SIEDS)*, pp. 117–121, Charlottesville, VA, USA, April 2017.
- [21] N. S. Halvaiee and M. K. Akbari, "A novel model for credit card fraud detection using artificial immune systems," *Applied Soft Computing*, vol. 24, pp. 40–49, 2014.
- [22] J. T. S. Quah and M. Sriganesh, "Real-time credit card fraud detection using computational intelligence," *Expert Systems with Applications*, vol. 35, no. 4, pp. 1721–1732, 2008.
- [23] F. Carcillo, Y.-A. L. Borgne, O. Caelen, and G. Bontempi, "An assessment of streaming active learning strategies for real-life credit card fraud detection," in *2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pp. 631–639, Tokyo, TKY, Japan, October 2017.
- [24] X. Zhang, Y. Han, W. Xu, and Q. Wang, "HOBA: a novel feature engineering methodology for credit card fraud detection with a deep learning architecture," *Information Sciences*, vol. 557, pp. 302–316, 2021.
- [25] P. Cao, X. Liu, J. Zhang, D. Zhao, M. Huang, and O. Zaiane, " $\ell_{2,1}$  norm regularized multi-kernel based joint nonlinear feature selection and over-sampling for imbalanced data classification," *Neurocomputing*, vol. 234, pp. 38–57, 2017.
- [26] F. Zhang, G. Liu, Z. Li, C. Yan, and C. Jiang, "GMM-based undersampling and its application for credit card fraud detection," in *2019 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, Budapest, BP, Hungary, July 2019.
- [27] H. He, W. Zhang, and S. Zhang, "A novel ensemble method for credit scoring: adaption of different imbalance ratios," *Expert Systems with Applications*, vol. 98, pp. 105–117, 2018.
- [28] F. Thabtah, S. Hammoud, F. Kamalov, and A. Gonsalves, "Data imbalance in classification: experimental evaluation," *Information Sciences*, vol. 513, pp. 429–441, 2020.
- [29] A. Singh and A. Jain, "An empirical study of AML approach for credit card fraud detection-financial transactions," *International Journal of Computers Communications & Control*, vol. 14, no. 6, pp. 670–690, 2020.
- [30] A. Armel and D. Zaidouni, "Fraud detection using apache spark," in *2019 5th International Conference on Optimization and Applications (ICOA)*, pp. 1–6, Kenitra, Morocco, April 2019.
- [31] E. Kim, J. Lee, H. Shin et al., "Champion-challenger analysis for credit card fraud detection: Hybrid ensemble and deep learning," *Expert Systems with Applications*, vol. 128, pp. 214–224, 2019.
- [32] Z. Li, G. Liu, and C. Jiang, "Deep representation learning with full center loss for credit card fraud detection," *IEEE Transactions on Computational Social Systems*, vol. 7, no. 2, pp. 569–579, 2020.
- [33] Y. Lucas, P.-E. Portier, L. Laporte et al., "Towards automated feature engineering for credit card fraud detection using multi-perspective HMMs," *Future Generation Computer Systems*, vol. 102, pp. 393–402, 2020.
- [34] Y. Wu, Y. Xu, and J. Li, "Feature construction for fraudulent credit card cash-out detection," *Decision Support Systems*, vol. 127, article 113155, 2019.
- [35] V. V. Vlasselaer, C. Bravo, O. Caelen et al., "APATE: a novel approach for automated credit card transaction fraud detection using network-based extensions," *Decision Support Systems*, vol. 75, pp. 38–48, 2015.

- [36] G. Gianini, L. G. Fossi, C. Mio, O. Caelen, L. Brunie, and E. Damiani, "Managing a pool of rules for credit card fraud detection by a game theory based approach," *Future Generation Computer Systems*, vol. 102, pp. 549–561, 2020.
- [37] U. Fiore, A. D. Santis, F. Perla, P. Zanetti, and F. Palmieri, "Using generative adversarial networks for improving classification effectiveness in credit card fraud detection," *Information Sciences*, vol. 479, pp. 448–455, 2019.
- [38] F. Carcillo, A. D. Pozzolo, Y.-A. L. Borgne, O. Caelen, Y. Mazzer, and G. Bontempi, "SCARFF: a scalable framework for streaming credit card fraud detection with spark," *Information Fusion*, vol. 41, pp. 182–194, 2018.
- [39] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, no. 1, pp. 321–357, 2002.
- [40] H. Han, W.-Y. Wang, and B.-H. Mao, "Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning," in *Advances in Intelligent Computing, International Conference on Intelligent Computing (ICIC)*, pp. 878–887, Hefei, China, August 2005.
- [41] Q. Fan, Z. Wang, and D. Gao, "One-sided dynamic undersampling no-propagation neural networks for imbalance problem," *Engineering Applications of Artificial Intelligence*, vol. 53, pp. 62–73, 2016.
- [42] P. Vuttipittayamongkol and E. Elyan, "Neighbourhood-based undersampling approach for handling imbalanced and overlapped data," *Information Sciences*, vol. 509, pp. 47–70, 2020.
- [43] A. Iranmehr, H. Masnadi-Shirazi, and N. Vasconcelos, "Cost-sensitive support vector machines," *Neurocomputing*, vol. 343, pp. 50–64, 2019.
- [44] A. C. Bahnsen, D. Aouada, and B. Ottersten, "Example-dependent cost-sensitive logistic regression for credit scoring," in *13th International Conference on Machine Learning and Applications (ICMLA)*, pp. 263–269, Detroit, MI, USA, December 2014.
- [45] S. Akila and U. S. Reddy, "Cost-sensitive risk induced Bayesian inference bagging (RIBIB) for credit card fraud detection," *Journal of Computational Science*, vol. 27, pp. 247–254, 2018.
- [46] S. Nami and M. Shajari, "Cost-sensitive payment card fraud detection based on dynamic random forest and  $k$ -nearest neighbors," *Expert Systems with Applications*, vol. 110, pp. 381–392, 2018.
- [47] Kaggle, "Kaggle," November 2020, <https://www.kaggle.com/mlg-ulb/creditcardfraud>.
- [48] A. C. Bahnsen, D. Aouada, A. Stojanovic, and B. Ottersten, "Feature engineering strategies for credit card fraud detection," *Expert Systems with Applications*, vol. 51, pp. 134–142, 2016.

## Research Article

# Angle Estimation Using Local Searching for Bistatic MIMO Radar with Unknown MCM

Chaochen Tang <sup>1,2</sup> Hongbing Qiu <sup>3</sup> Xin Liu <sup>2</sup> and Qinghua Tang <sup>2</sup>

<sup>1</sup>School of Telecommunications Engineering, Xidian University, Xi'an, Shaanxi 710071, China

<sup>2</sup>School of Information Science and Engineering, Guilin University of Technology, Guilin, Guangxi 541004, China

<sup>3</sup>School of Information and Communication, Guilin University of Electronic Technology, Guilin, Guangxi 541004, China

Correspondence should be addressed to Chaochen Tang; [gxtcc2008@126.com](mailto:gxtcc2008@126.com)

Received 12 March 2021; Revised 11 June 2021; Accepted 1 July 2021; Published 16 July 2021

Academic Editor: Shan Zhong

Copyright © 2021 Chaochen Tang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Multiple input and multiple output (MIMO) radar systems have advantages over traditional phased-array radar in resolution, parameter identifiability, and target detection. However, the estimation performance of the direction of arrivals (DOAs) and the direction of departures (DODs) will be significantly degraded for a colocated MIMO radar system with unknown mutual coupling matrix (MCM). Although auxiliary sensors (AS) can be set to solve this problem, the computational cost of two-dimensional multiple signal classification (2D-MUSIC) is still large. In this paper, a new angle estimation method is proposed to reduce the computational complexity. First, a local-search range is defined for each initial angle estimation obtained by the MUSIC with AS method. Second, the new estimation of DOAs and DODs of the targets is estimated via the joint estimation theory of angle and mutual coupling coefficient in the local search area. Simulation results validate that the proposed method can obtain the same precision and have the advantage over the global searching in computational complexity.

## 1. Introduction

A MIMO radar system transmits orthogonal waveforms via its antennas, which can supply more independent transmit/receive channels than that of traditional phased-array radar. There are two types of MIMO radar systems: one is statistical MIMO radar [1, 2] and the other is colocated MIMO radar [3]. The former is composed of widely separated antennas, which can achieve the spatial diversity gain and overcome the scintillation effect of targets. The antenna configuration of the latter is the same as that of phased array radar and can form virtual arrays, which result in performance improvement of target detection and parameter identification. In our work, we focus on a bistatic radar system based on the latter type of MIMO radar system.

The bistatic MIMO radar system combines the advantages of MIMO radar and bistatic radar system, which has been researched extensively. How to improve the estimation performance of direction-of-arrivals (DOAs) and direction of departures (DODs) is one of the hot research issues in existing studies. A great number of estimation methods have

been proposed, such as maximum likelihood (ML) [4–7], subspace-based [8–12], and sparse signal representation (SSR) [13–15]. Compared to ML and SSR, multiple signal classification (MUSIC) and estimation of signal parameters via rotational invariance technique (ESPRIT) are two typical methods which can achieve high estimation performance with low computational complexity than that of ML and SSR [16]. These methods will work well under ideal conditions. In a real scenario, however, there is a mutual coupling effect between the elements of array, which has a great influence on the performance of angle estimation methods [17–20], and there is no exception for co-located MIMO radar using arrays [21–23].

There are many methods for solving the problem of mutual coupling effect, such as self-calibration method [24], sparse array method [25], and auxiliary sensors (AS) method [17, 21, 26, 27]. Among these methods, the AS method is proved to be the most effective and feasible method [17, 21, 26, 27]. In this method, the first and the last sensors of the transmit and receive array are set as auxiliary sensors. After that, the effects of unknown MCM can be eliminated, and then the MUSIC method can be directly applied to estimate

the DOAs and DODs of targets. Unfortunately, AS method results in the loss of array aperture, which is caused by the reduction of the number of available array elements [17, 28, 29]. More importantly, there is high computational complexity since two-dimensional global searching to be performed by MUSIC in the process of the angle estimation [30–32].

In this paper, a local searching algorithm for the estimation of DOAs and DODs for the colocated MIMO radar in the presence of mutual coupling is proposed. The method has three phases. First, initial DOAs and DODs are estimated via AS method without knowing the mutual coupling coefficients. Second, a local search range is defined for each pair of DOAs and DODs obtained in the first phase. Finally, new angle estimates are obtained based on the theory of joint angle and mutual coupling coefficient estimation. We make a lot of analysis on the key factors, such as search range and search step length, which are closely related to the computational complexity and estimation accuracy of our algorithm. Simulation results validate that the proposed algorithm can achieve good performance in terms of lower complexity and high estimation accuracy.

This paper is organized as follows. Section 2 introduces the signal model of colocated MIMO system. Section 3 presents the process of the local search method for angle estimation and theoretically analyzes the computational complexity of that. Simulation results are given in Section 4 to demonstrate the performance of the proposed method. And Section 5 is the conclusion.

## 2. Signal Model

Consider a bistatic MIMO radar system with a transmitter and a receiver. Both the transmitter and the receiver are uniform linear arrays (ULA), which are equipped with  $M$  transmitting antennas and  $N$  receiving antennas, respectively. The distance between adjacent antennas is half a wavelength. It is assumed that there are  $Q$  targets in far-field, the DOD and DOA of the  $q$ th target relative to the transmitter and the receiver are  $\varphi_q$  and  $\theta_q$ , respectively. In a MIMO radar system, the transmitter transmits  $M$  orthogonal waveforms  $\mathbf{S}_m(t)$ ,  $m = 1, 2, \dots, M$ , which satisfies the condition of  $\int_{T_p} \mathbf{S}(t)\mathbf{S}^H(t) = \mathbf{I}_M$ . Here,  $T_p$  is the duration of the pulse, and  $\mathbf{I}$  is the identity matrix. Thus, the received data at time  $t$  in the  $k$ th snapshot can be written as

$$\mathbf{x}_k(t) = \sum_{q=1}^Q \beta_{qk} \mathbf{a}(\theta_q) \mathbf{b}^T(\varphi_q) \mathbf{S}(t) + \mathbf{n}_k(t), \quad (1)$$

where  $\mathbf{x}_k(t) \in \mathbb{C}^{N \times 1}$  is the received data vector,  $\beta_{qk}$  is the channel parameter representing reflection coefficient and Doppler frequency,  $\mathbf{a}(\theta_q) = [1, e^{j\pi \sin \theta_q}, \dots, e^{j\pi(M-1) \sin \theta_q}]^T$  and  $\mathbf{b}(\theta_q) = [1, e^{j\pi \sin \varphi_q}, \dots, e^{j\pi(N-1) \sin \varphi_q}]^T$  are receive steering vector and transmitting steering vector, respectively,  $\mathbf{S}(t) = [s_1(t), s_2(t), \dots, s_M(t)]^T$  is the signal vector, and  $\mathbf{n}_k(t)$  is considered as the unknown uniform noise.

After matched filtering by  $M$  matched filters, a matrix form of the output data is given by

$$\mathbf{X}_k = \mathbf{A}(\theta) \boldsymbol{\Sigma}_k \mathbf{B}^T(\varphi) + \mathbf{N}_k, \quad (2)$$

where  $\mathbf{X}_k \in \mathbb{C}^{M \times N}$  is the output data,  $\mathbf{A}(\theta) = [\mathbf{a}(\theta_1), \mathbf{a}(\theta_2), \dots, \mathbf{a}(\theta_Q)]$  and  $\mathbf{B}(\theta) = [\mathbf{b}(\varphi_1), \mathbf{b}(\varphi_2), \dots, \mathbf{b}(\varphi_Q)]$  are the transmit steering matrix and the receiving steering matrix, respectively,  $\boldsymbol{\Sigma}_k = \text{diag} \{ \beta_{1k}, \beta_{1k}, \dots, \beta_{Qk} \}$  is the channel parameters matrix, and  $\mathbf{N}_k \in \mathbb{C}^{M \times N}$  is the noise.

However, the mutual coupling effect exists objectively in the arrays, which is not reflected in formula (2). In general, they can be described as mutual coupling matrix (MCM), also called band-symmetric Toeplitz matrix. So, in the case of mutual coupling, the actual data that the receiver outputs should be expressed as

$$\mathbf{X}_k = [\mathbf{C}_t \mathbf{A}(\theta)] \boldsymbol{\Sigma}_k [\mathbf{C}_r \mathbf{B}(\varphi)]^T + \mathbf{N}_k, \quad (3)$$

where  $\mathbf{C}_t = \text{toeplitz}([1, c_1, 0, \dots, 0]) \in \mathbb{C}^{M \times M}$  and  $\mathbf{C}_r = \text{toeplitz}([1, c_2, 0, \dots, 0]) \in \mathbb{C}^{N \times N}$  represent the MCM of transmit arrays and receive arrays, respectively.

In order to process MIMO radar data, it is necessary to change the matrix  $\mathbf{X}_k$  into a column vector  $\mathbf{Y}_k$ , which is given by

$$\mathbf{Y}_k = \text{vec}(\mathbf{X}_k) = [\mathbf{C}_t \otimes \mathbf{C}_r] [\mathbf{A}(\theta) \odot \mathbf{B}(\varphi)] \boldsymbol{\beta}_k + \mathbf{W}_k = \mathbf{C}\mathbf{K}(\theta, \varphi) \boldsymbol{\beta}_k + \mathbf{W}_k, \quad (4)$$

where  $\text{vec}(\cdot)$ ,  $\mathbf{C} = \mathbf{C}_t \otimes \mathbf{C}_r$ , and  $\mathbf{A}(\theta) \odot \mathbf{B}(\varphi) = \mathbf{K}(\theta, \varphi)$  denote the vectorization operation, the equivalent of MCM, and the virtual steering matrix, respectively. The symbols  $\otimes$  and  $\odot$  stand for Kronecker product and Hadamard product, respectively.  $\mathbf{K}(\theta, \varphi) = [\mathbf{k}(\theta_1, \varphi_1), \dots, \mathbf{k}(\theta_Q, \varphi_Q)]$ , in which the  $\mathbf{k}(\theta_q, \varphi_q)$  satisfied with  $\mathbf{k}(\theta_q, \varphi_q) = \mathbf{a}(\theta_q) \otimes \mathbf{b}(\varphi_q)$ .  $\boldsymbol{\beta}_k = \text{vec}(\boldsymbol{\Sigma}_k)$ , and  $\mathbf{W}_k = \text{vec}(\mathbf{N}_k)$ .

The sampling covariance matrix of  $\mathbf{Y}_k$  can be written as:

$$\mathbf{R}_Y = E\{\mathbf{Y}_k \mathbf{Y}_k^H\} = \mathbf{C}\mathbf{K}(\theta, \varphi) \boldsymbol{\beta}_k \boldsymbol{\beta}_k^H [\mathbf{C}\mathbf{K}(\theta, \varphi)]^H + \sigma^2 \mathbf{I}_{MN}, \quad (5)$$

where  $\sigma^2$  is the average noise power. If  $\mathbf{C}$  is known, then the noise subspace and the signal subspace can be obtained by Eigenvalue decomposition of equation (5). Furthermore, the DOA can be estimated through subspace method. The truth of the matter that  $\mathbf{C}$  is usually unknown, which will result in errors in angle estimation.

## 3. Proposed Method

**3.1. Method.** The proposed method consists of three phases as follows.

**3.1.1. Phase 1.** In the first phase, the sensors on either sides of the transmitter/receiver array are set to be auxiliary sensors, which are marked with grey color and shown in Figure 1. Then, we use the algorithm in reference [29] to obtain the initial angle estimations of all targets. It is noted that the

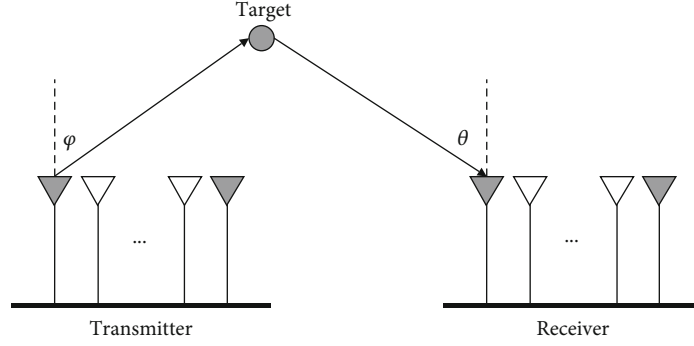


FIGURE 1: The single base MIMO radar array model with auxiliary sensors.

search step length can be allowed to be set longer in the process of spectral peak search. For example, the step length is 0.1 degree instead of 0.01 degree. Although the precision becomes worse with longer step length, the computational cost is significantly reduced in the same search area.

**3.1.2. Phase 2.** In order to compensate for the decrease of accuracy caused by the first phase,  $Q$  local search areas with the same size are defined for  $Q$  targets. And we let the initial estimation of  $q$ th target  $\langle \tilde{\theta}_q, \tilde{\varphi}_q \rangle$  from phase 1 as the center of the  $q$ th local search area  $\langle \tilde{\theta}_q \pm \Delta\theta, \tilde{\varphi}_q \pm \Delta\varphi \rangle$ . The details are shown in Figure 2. In Figure 2,  $\Delta\theta$  and  $\Delta\varphi$  indicate the size of the local search area. And the red and blue dots represent one of the initial angle estimation value obtained in phase 1 and the angle estimations to be estimated, respectively. To simplify, we assumed that  $\Delta\theta = \Delta\varphi$  in this paper.

**3.1.3. Phase 3.** In the third phase, we define  $(S + 1)$  points of the DOAs and the DODs to be searched in the  $q$ th local search area for the  $q$ th targets are  $\langle \tilde{\theta}_q - \Delta\theta, \dots, \tilde{\theta}_q - \theta_s, \tilde{\theta}_q, \tilde{\theta}_q - \theta_s, \dots, \tilde{\theta}_q + \Delta\theta \rangle$  and  $\langle \tilde{\varphi}_q - \Delta\varphi, \dots, \tilde{\varphi}_q - \varphi_s, \tilde{\varphi}_q, \tilde{\varphi}_q - \varphi_s, \dots, \tilde{\varphi}_q + \Delta\varphi \rangle$ , respectively, where  $\theta_s = (2\Delta\theta)/S$ . Then, the total number of search points in the  $q$ th local search are  $(S + 1)^2$ . Indeed, this phase will increase the cost of computation, but its computation is still much less than that of MUSIC with search step 0.01 degree. We will discuss it in detail later in the section of simulation. After completing the above procedure, the final angle estimation of all targets is based on the theory of joint estimation of angle and mutual coupling coefficient [17].

In equation (4), the mutual coupling matrix can be expressed as

$$\mathbf{C} = \mathbf{C}_t \otimes \mathbf{C}_r = \begin{bmatrix} \mathbf{C}_1 & \mathbf{C}_2 & 0 & \cdots & 0 \\ \mathbf{C}_2 & \mathbf{C}_1 & \mathbf{C}_2 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \mathbf{C}_2 & \mathbf{C}_1 & \mathbf{C}_2 \\ 0 & \cdots & 0 & \mathbf{C}_2 & \mathbf{C}_1 \end{bmatrix}, \quad (6)$$

where  $\mathbf{C}_1$  and  $\mathbf{C}_2$  are Toeplitz matrix, which are given by

$$\begin{cases} \mathbf{C}_1 = \mathbf{C}_r = \text{toeplitz}\{[1, c_2, 0, \dots, 0]\} \in \mathbb{C}^{N \times N}, \\ \mathbf{C}_2 = c_1 \cdot \mathbf{C}_r = \text{toeplitz}\{[c_1, c_1 c_2, 0, \dots, 0]\} \in \mathbb{C}^{N \times N}, \end{cases} \quad (7)$$

where  $c_1$  and  $c_2$  are the unknown mutual coupling coefficients of transmit arrays and receive arrays, respectively.

According to the literature [17], we have

$$\mathbf{C}_k \mathbf{a}(\theta, \rho) = \mathbf{T}_N [\mathbf{a}(\theta)] \rho_k, \quad k = 1, 2, \quad (8)$$

where  $\rho_1 = [1, c_1]^T$ ,  $\rho_2 = [c_1, c_1 c_2]^T$ ,  $\mathbf{T}_N = \mathbf{T}_N^1 + \mathbf{T}_N^2$  is a transformation matrix which can be given by

$$\begin{cases} [\mathbf{T}_N^1]_{i,j} = \begin{cases} \mathbf{a}(\theta)_{i+j-1} & i + j \leq N + 1, \\ 0 & \text{otherwise,} \end{cases} \\ [\mathbf{T}_N^2]_{i,j} = \begin{cases} \mathbf{a}(\theta)_{i-j+1} & i \geq j \geq 2, \\ 0 & \text{otherwise.} \end{cases} \end{cases} \quad (9)$$

A transformation matrix  $\mathbf{T}_M = \mathbf{T}_M^1 + \mathbf{T}_M^2$  is similarly defined as

$$[\mathbf{T}_M^1]_{i,j} = \begin{cases} \mathbf{b}(\varphi)_{i+j-1} & i + j \leq M + 1, \\ 0 & \text{otherwise,} \end{cases} \quad (10)$$

$$[\mathbf{T}_M^2]_{i,j} = \begin{cases} \mathbf{b}(\varphi)_{i-j+1} & i \geq j \geq 2, \\ 0 & \text{otherwise.} \end{cases} \quad (11)$$

From equations (8)–(11), we have

$$\mathbf{C} \mathbf{k}(\varphi, \theta) = \mathbf{C} [\mathbf{b}(\varphi) \otimes \mathbf{a}(\theta)] = (\mathbf{T}_M \otimes \mathbf{T}_N) \begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix} = \mathbf{T}(\varphi, \theta) \rho, \quad (12)$$

where  $\rho = [1, c_1, c_1, c_1 c_2]^T$ . The estimates of DOAs can be obtained by the following equation



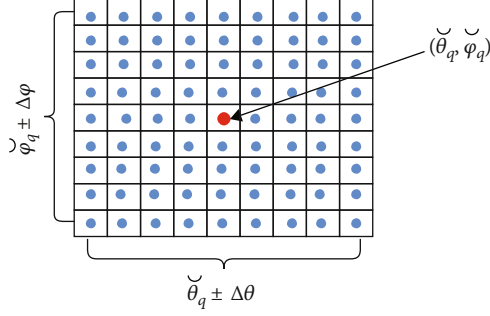


FIGURE 2: A local search area around the initial estimation of the  $q$ th target.

$$\left\| \mathbf{E}_n^H \mathbf{C} \mathbf{k}(\varphi_q, \theta_q) \right\|^2 = \left\| \mathbf{E}_n^H \mathbf{T}(\varphi_q, \theta_q) \rho \right\|^2 = 0, q = 1, 2, \dots, Q. \quad (13)$$

The equivalent expression of the equation (13) is as follows

$$\rho^H \mathbf{T}^H \mathbf{E}_n \mathbf{E}_n^H \mathbf{T} \rho = 0. \quad (14)$$

In formula (14), the vector  $\rho$  is regarded as an eigenvector, which can be obtained via the eigen decomposition of  $\mathbf{T}^H \mathbf{E}_n \mathbf{E}_n^H \mathbf{T}$ . And the eigenvalue is zero when the estimates of the angle are equal to the true DOAs. However, the estimates always deviate from the real values. Therefore, an optimization function is constructed, which is given by

$$\langle \hat{\varphi}, \hat{\theta} \rangle = \arg \min_{\varphi, \theta} (\lambda_{\min}[\mathbf{D}(\varphi, \theta)]), \quad (15)$$

where  $\mathbf{D}(\varphi, \theta) = \mathbf{T}^H \mathbf{E}_n \mathbf{E}_n^H \mathbf{T}$ , and the function of  $\lambda_{\min}[\cdot]$  is to select the smallest one from the eigenvalues via ED of  $\mathbf{D}(\varphi, \theta)$ ,  $i = 1, 2, \dots, N^2 + 1$ . In other words, for a local search range, where the angles with minimum eigenvalue are new estimates of DOAs. The eigenvector  $\rho$  is corresponding to the minimum eigenvalue, and the mutual coupling coefficients matrix  $\rho$  is corresponding to the minimum eigenvalue.

$$\rho = e_{\min}[\mathbf{D}(\tilde{\theta}, \tilde{\varphi})]. \quad (16)$$

In order to better demonstrate our method, we summarize the realization process of the proposed algorithm in Algorithm 1.

**3.2. Computational Complexity of Proposed Method.** The value of step length is usually set as  $0.01^\circ$  in global searching. That means the number of search points is 18001 in the range of  $-90^\circ \sim 90^\circ$ . The computational cost of  $\mathbf{T}^H \mathbf{E}_n \mathbf{E}_n^H \mathbf{T}$  is  $O[8MN(MN - Q) + 16MN]$ , and the complexity of eigen decomposition of  $\mathbf{D}(\varphi, \theta) \in \mathbb{C}^{4 \times 4}$  is  $O(64)$  for each point via joint estimation of angles and mutual coupling coefficients method. Therefore, the total computational cost of the global searching is  $O\{18001 * 64 * [8MN(MN - Q) + 16MN]\}$ .

Different from the global searching method, the main calculation process of the proposed method consists of two parts. The first part is to achieve the initial angle estimation via MUISC with the auxiliary element method. The computational cost of this part is  $O[3601 * MN(MN - Q) + (MN - Q)]$  under the assumption that the search step length is  $0.05^\circ$  and the search range is  $-90^\circ \sim 90^\circ$ . The angle estimation method of the second part is the same as the first part but with local search, in which the computational cost is  $O\{Q(S + 1)^2 * 64 * [8MN(MN - Q) + 16MN]\}$ . The complexity of the local search method is  $O\{[3601 * MN(MN - Q) + (MN - Q)] + Q(S + 1)^2 * 644[8MN(MN - Q) + 16MN]\}$ .

Let  $S = 10$  and set the variation range of the local search to be  $\pm 0.05^\circ$ , then the local search method can achieve the same angle estimation accuracy as the global search method. For simplicity, let  $Q = MN - 1$ . To compare the computational complexity of the two methods, we define the computational cost ratio of the global search method to the local search method as  $\gamma$ . When  $M = N = 12, 10, 8$ ,  $\gamma$  is equal to 1.0402, 1.5024, and 2.2367, respectively. It is indicated that the local search method has the advantage over the global search in lightweight applications.

## 4. Simulation Results and Discussions

In this section, we will demonstrate the performance of the proposed method via three simulations. Consider a L-type bistatic MIMO radar whose transmitter and receiver are ULAs with 8 array elements, which are separated by half wavelength. And  $Q = 3$  known uncorrelated targets are located in the far-field with angles  $(\theta_1 = 10^\circ, \varphi_1 = 10^\circ)$ ,  $(\theta_2 = 30^\circ, \varphi_2 = 30^\circ)$ , and  $(\theta_3 = 45^\circ, \varphi_3 = 45^\circ)$ , respectively. The transmit signals are orthogonal discrete multifrequency signals  $[*]$ , and additive noise is zero mean, i.i.d., white Gaussian processes with variance  $\sigma^2$ . We set the number of Monte Carlo trial ( $K$ ) and snapshots ( $P$ ) to be 500 and 100, respectively. The validity of angle estimation is measured by root mean square error (RMSE):

$$RMSE = \sqrt{\frac{1}{2KQ} \sum_{k=1}^K \sum_{q=1}^Q [(\theta_{k,q} - \theta_q)^2 + (\varphi_{k,q} - \varphi_q)^2]}, \quad (17)$$

where  $\langle \hat{\theta}_{k,q}, \hat{\varphi}_{k,q} \rangle$  denotes the  $q$ th DOA and DOD in the  $k$ th trial, and  $\langle \theta_q, \varphi_q \rangle$  denotes the true DOA and DOD of the  $q$ th target.

**4.1. Simulation Experiment 1: RMSE Simulation of Angle Estimation of the Proposed Method with Different Segments.** Before doing this experiment, we had obtained the initial angle estimations  $\{\tilde{\theta}_q, \tilde{\varphi}_q\}$  of  $Q$  known targets via auxiliary elements method, in which the search step length was set to be  $0.1^\circ$ . In this experiment, the local search range of each angle estimation is  $\langle \tilde{\theta}_q \pm 0.5^\circ, \tilde{\varphi}_q \pm 0.5^\circ \rangle$ .  $S_1 = 10$ ,  $S_2 = 20$ , and  $S_3 = 40$  are different segments, and the corresponding

- 1: Use MUSIC with AS method, find the initial angle estimation  $\langle \tilde{\theta}_q, \tilde{\varphi}_q \rangle$ .
- 2: Define a local search range  $\langle \tilde{\theta}_q \pm \Delta\theta, \tilde{\varphi}_q \pm \Delta\varphi \rangle$ , divide  $\tilde{\theta}_q \pm \Delta\theta$  and  $\tilde{\varphi}_q \pm \Delta\varphi$  into  $S+1$  segments, obtain  $(S+1)^2$  angle points to be estimated.
- 3: Compute the mutual coupling matrix  $C$  and the transformation matrix  $T$  according to (6) and (8)–(12), respectively.
- 4: Using noise covariance matrix  $E_n$  and  $T$  obtained in step 3, compute  $D(\varphi, \theta)$  according to  $D(\varphi, \theta) = T^H E_n T$  for each angle point.
- 5: Compute the eigen decomposition of  $D(\varphi, \theta)$ .
- 6: The angle point corresponding to the minimal eigenvalue obtained from step 5, which is select as angle estimation of the  $q$ th target according to (15).
- 7: Obtain the mutual coupling coefficients according to (16).

ALGORITHM 1: The proposed method.

search step length is  $0.1^\circ$ ,  $0.05^\circ$ , and  $0.025^\circ$ , respectively. The number of points to be estimated is 121, 441, and 1681, respectively.

Figure 3 shows the RMSEs of the proposed method with different segments in the case of the fixed search range. From Figure 3, we can see that the proposed method can achieve better performance with the increase in the number of search segments.

**4.2. Simulation Experiment 2: RMSE Simulation of Angle Estimation of the Proposed Method with Different Search Range.** For simplicity, the number of search segments is fixed at 10 in this experiment. That is to say, the number of search points is 121. By setting the search step length  $d$  to be  $0.05^\circ$  and  $0.1^\circ$ , we get the search range is  $\pm 0.25^\circ$  and  $\pm 0.5^\circ$ , respectively. Figure 4 shows the RMSEs of the proposed method with different search range in the case of fixed search segments. We have noted that large search step length provides better performance in the case of lower SNR. This is because there is a large deviation between true angle and initial angle estimation obtained under the conditions of low SNR. As a result, the probability of a small search range covering the true angle is less than that of a large search range. With the increase of SNR, the performance of small range search is improved significantly.

**4.3. Simulation Experiment 3: Comparison of the Proposed Method, Standard MUSIC, and MUSIC with Auxiliary Sensors.** Experiments 1 and 2 merely show that the proposed method itself is affected by the search range and search step length. To prove the good performance of the proposed method under the condition of unknown MCM, we make a comparison of it with traditional DOA estimation methods based on spectral peak search, such as standard MUSIC method and MUSIC with auxiliary sensors method. At first, we set the search length to be  $0.1^\circ$  for all methods tested. Then, for the proposed method, a local search is established with a range from  $-0.05^\circ$  to  $+0.05^\circ$ , and segments are equal to 11, which resulting in 121 grid points is necessary to be estimated. Figure 5 shows the estimation RMSEs of the three different methods. From Figure 5, we can see that standard MUSIC cannot estimate the angles under the condition of mutual coupling. The MUSIC with auxiliary sensors method can obtain the angle estimation, but its RMSE is larger than that of the proposed method. The same RMSE will be

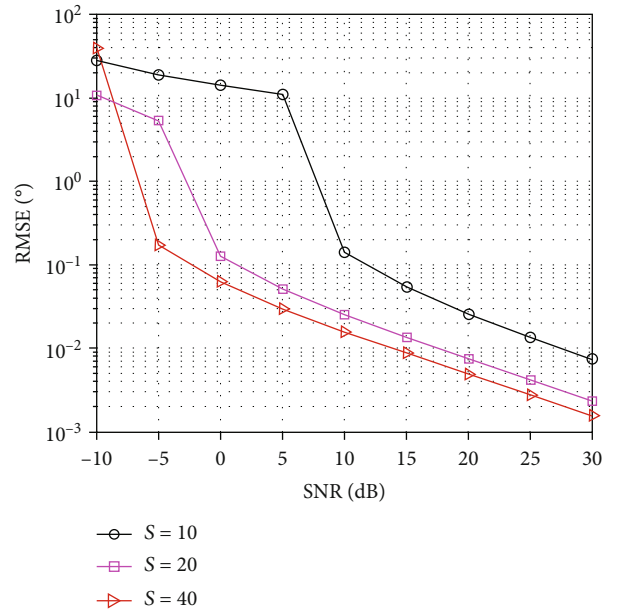
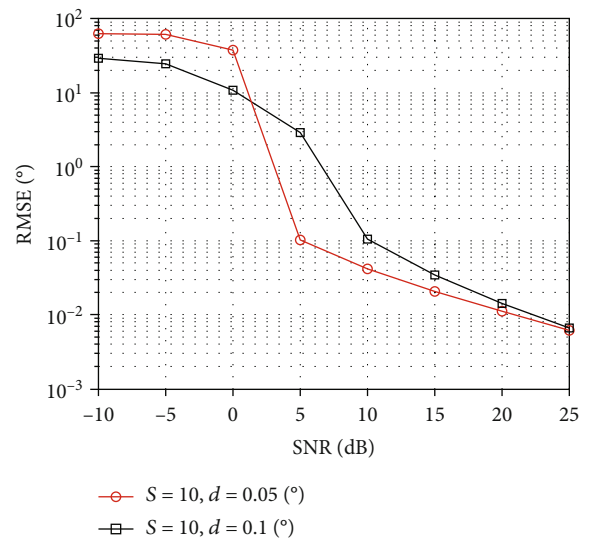


FIGURE 3: The RMSEs versus SNR for segments = 10, 20, and 40.

FIGURE 4: The RMSEs versus SNR for search step length  $\pm 0.25^\circ$  and  $\pm 0.5^\circ$ .

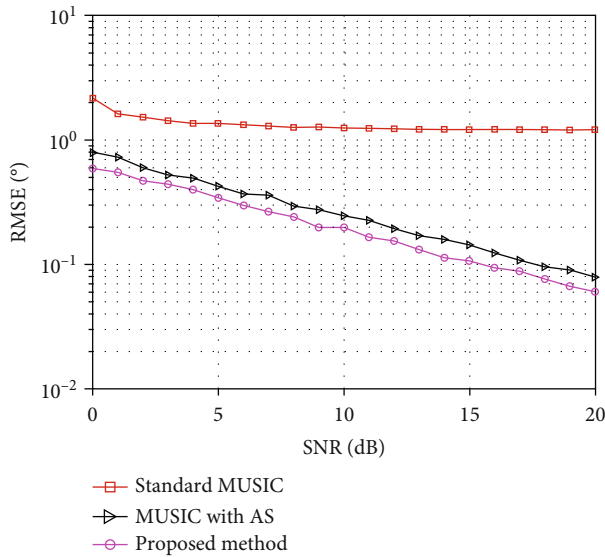


FIGURE 5: The RMSEs of angle estimation versus SNR in experiment 3.

obtained if the search step length is equal. But the MUSIC with auxiliary sensors has a higher computational complexity than that of our method, which is proved in Section 3.

## 5. Conclusion

In this paper, we have studied the problem of DOAs and DOD estimation of bistatic MIMO radar in the presence of unknown mutual coupling. The auxiliary sensor method is one of the effective methods to solve this problem. But it still has a high computational cost while the MUSIC algorithm is used for 2D global searching. To reduce the computational cost of that, a local search method has been proposed. The angle estimation performance of the proposed method has an advantage over that of MUSIC with auxiliary element method via global searching in computational complexity. The main reason is that it allows the auxiliary element method to obtain the initial value of the angle with a larger search step length. And the angular accuracy degradation, which caused by globe searching with large search length, can be compensated by local searching. In addition, the mutual coupling coefficients can be obtained based on the joint estimation theory. Our work shows that the proposed algorithm is more suitable for lightweight applications.

## Data Availability

The data included in this paper are available without any restriction.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

- [1] E. Fishler, A. Haimovich, R. Blum, D. Chizhik, L. Cimini, and R. Valenzuela, "MIMO radar: an idea whose time has come," in *Proceedings of the 2004 IEEE Radar Conference (IEEE Cat. No.04CH37509)*, pp. 71–78, Philadelphia, PA, USA, 2004.
- [2] E. Fishler, A. Haimovich, R. S. Blum, L. J. Cimini, D. Chizhik, and R. A. Valenzuela, "Spatial diversity in radars-models and detection performance," *IEEE Transactions on Signal Processing*, vol. 54, no. 3, pp. 823–838, 2006.
- [3] D. W. Bliss and K. W. Forsythe, "Multiple-input multiple-output (MIMO) radar and imaging: degrees of freedom and resolution," in *The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, vol. 1, pp. 54–59, Pacific Grove, CA, USA, 2003.
- [4] B. Tang, J. Tang, Y. Zhang, and Z. Zheng, "Maximum likelihood estimation of DOD and DOA for bistatic MIMO radar," *Signal Processing*, vol. 93, no. 5, pp. 1349–1357, 2013.
- [5] J. Paik and J. Lee, "Performance analysis of ML-based DOA estimation algorithm in bistatic MIMO radar system," in *2017 7th IEEE International Symposium on Microwave, Antenna, Propagation, and EMC Technologies (MAPE)*, pp. 543–548, Xi'an, China, 2017.
- [6] H. Chen, D. Yang, H. Wang, X. Li, and Z. Zhuang, "Direction finding for bistatic mimo radar using Em maximum likelihood algorithm," *Progress In Electromagnetics Research*, vol. 141, pp. 99–116, 2013.
- [7] J. Gong and Y. Guo, "A bistatic MIMO radar angle estimation method for coherent sources in impulse noise background," *Wireless Personal Communications*, vol. 116, no. 1, pp. 3567–3576, 2020.
- [8] L. Li, F. Chen, and J. Dai, "Separate DOD and DOA estimation for bistatic MIMO radar," *International Journal of Antennas and Propagation*, vol. 2016, Article ID 9170403, 11 pages, 2016.
- [9] M. Bencheikh and Y. Wang, "Joint DOD-DOA estimation using combined ESPRIT-MUSIC approach in MIMO radar," *Electronics Letters*, vol. 46, no. 15, pp. 1081–1083, 2013.
- [10] M. Jin, G. Liao, and J. Li, "Joint DOD and DOA estimation for bistatic MIMO radar," *Signal Processing*, vol. 89, no. 2, pp. 244–251, 2009.
- [11] G. Zheng, B. Chen, and M. Yang, "Unitary ESPRIT algorithm for bistatic MIMO radar," *Electronics Letters*, vol. 48, no. 3, pp. 179–181, 2012.
- [12] G. Zheng, "Beamspace root-MUSIC algorithm for joint DOD DOA estimation in bistatic MIMO radar," *Wireless Personal Communications*, vol. 75, no. 4, pp. 1879–1889, 2014.
- [13] W. Tang, H. Jiang, and S. Pang, "Low complexity 3D-OMP algorithms for DOD DOA and Doppler frequency estimation in bistatic MIMO radar," *International Journal of Electronics*, vol. 106, no. 6, pp. 816–828, 2019.
- [14] Z. Cao, L. Zhou, and J. Dai, "Sparse Bayesian approach for DOD and DOA estimation with bistatic MIMO radar," *IEEE Access*, vol. 7, pp. 155335–155346, 2019.
- [15] Y. Lin and T. Lee, "Max-MUSIC: a low-complexity high-resolution direction finding method for sparse MIMO radars," *IEEE Sensors Journal*, vol. 20, no. 24, pp. 14914–14923, 2020.
- [16] M. Esfandiari, S. A. Vorobyov, S. Alibani, and M. Karimi, "Non-iterative subspace-based DOA estimation in the presence of nonuniform noise," *IEEE Signal Processing Letters*, vol. 26, no. 6, pp. 848–852, 2019.

- [17] Z. Ye and C. Liu, "2-D DOA estimation in the presence of mutual coupling," *IEEE Transactions on Antennas and Propagation*, vol. 56, no. 10, pp. 3150–3158, 2008.
- [18] B. Friedlander and A. J. Weiss, "Direction finding in the presence of mutual coupling," *IEEE Transactions on Antennas and Propagation*, vol. 39, no. 3, pp. 273–284, 1991.
- [19] X. Wang, G. Zhang, F. Wen, D. Ben, and W. Liu, "Angle estimation for bistatic MIMO radar with unknown mutual coupling based on three-way compressive sensing," *Journal of Systems Engineering and Electronics*, vol. 28, no. 2, pp. 257–266, 2017.
- [20] M. Liu, L. Zou, H. Ren, X. Yu, Y. Zhou, and X. Wang, "A novel MIMO array with reduced mutual coupling and increased degrees of freedom," *Mathematical Problems in Engineering*, vol. 2021, Article ID 3703657, 6 pages, 2021.
- [21] Z. Zheng, J. Zhang, and Y. Wu, "Multi-target localization for bistatic MIMO radar in the presence of unknown mutual coupling," *Journal of Systems Engineering and Electronics*, vol. 23, no. 5, pp. 708–714, 2012.
- [22] F. Wen, Z. Zhang, K. Wang, G. Sheng, and G. Zhang, "Angle estimation and mutual coupling self-calibration for ULA-based bistatic MIMO radar," *Signal Processing*, vol. 144, pp. 61–67, 2018.
- [23] Y. Guo, X. Wang, W. Wang et al., "Tensor-based angle estimation approach for strictly noncircular sources with unknown mutual coupling in bistatic MIMO radar," *Sensors*, vol. 18, no. 9, article 2788, 2018.
- [24] Y. Yue, Y. Xu, J. Zhuang, Y. Huang, K. Zhao, and Z. Liu, "Mutual coupling self-calibration for parameter estimation with vector antennas," in *2019 IEEE International Conference on Signal, Information and Data Processing (ICSIDP)*, pp. 1–5, Chongqing, China, 2019.
- [25] Z. Fu, P. Chargé, and Y. Wang, "Rearranged coprime array to increase degrees of freedom and reduce mutual coupling," *Signal Processing*, vol. 183, article 108038, 2021.
- [26] W. Si, D. Wu, L. Liu, and X. Qu, "Direction finding with gain/phase errors and mutual coupling errors in the presence of auxiliary sensors," *Mathematical Problems in Engineering*, vol. 2014, Article ID 429426, 12 pages, 2014.
- [27] Z. Zheng, Y. Yang, W. Wang, J. Yang, and Y. Ge, "2-D DOA estimation of multiple signals based on sparse L-shaped array," in *2016 International Symposium on Antennas and Propagation (ISAP)*, pp. 1014–1015, Okinawa, Japan, 2016.
- [28] X. Wang, W. Wang, J. Liu, Q. Liu, and B. Wang, "Tensor-based real-valued subspace approach for angle estimation in bistatic MIMO radar with unknown mutual coupling," *Signal Processing*, vol. 116, pp. 152–158, 2015.
- [29] Z. Zheng, J. Zhang, and J. Zhang, "Joint DOD and DOA estimation of bistatic MIMO radar in the presence of unknown mutual coupling," *Signal Processing*, vol. 92, no. 12, pp. 3039–3048, 2012.
- [30] X. Zhang, W. Chen, W. Zheng, Z. Xia, and Y. Wang, "Localization of near-field sources: a reduced-dimension MUSIC algorithm," *IEEE Communications Letters*, vol. 22, no. 7, pp. 1422–1425, 2018.
- [31] X. Zhang, L. Xu, L. Xu, and D. Xu, "Direction of departure (DOD) and direction of arrival (DOA) estimation in MIMO radar with reduced-dimension MUSIC," *IEEE Communications Letters*, vol. 14, no. 12, pp. 1161–1163, 2010.
- [32] J. Li, X. Zhang, R. Cao, and M. Zhou, "Reduced-dimension MUSIC for angle and array gain-phase error estimation in bistatic MIMO radar," *IEEE Communications Letters*, vol. 17, no. 3, pp. 443–446, 2013.

## Research Article

# Reader Scheduling for Tag Population Estimation in Multicategory and Multireader RFID Systems

Zhiyong He 

*College of Electrical Engineering, Nanjing Vocational University of Industry Technology, Nanjing 210046, China*

Correspondence should be addressed to Zhiyong He; 2012100740@niit.edu.cn

Received 9 May 2021; Accepted 25 June 2021; Published 12 July 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Zhiyong He. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Radio Frequency Identification (RFID) technology has been used in numerous applications, e.g., supply chain management and inventory control. This paper focuses on the practically important problem of the rapid estimation of the number of tags in large-scale RFID systems with multiple readers and multicategory RFID tags. RFID readers are often static and have to be deployed strategically after careful planning to cover the entire monitoring area, but reader-to-reader collision (R2Rc) remains a problem. R2Rc decreases the reliability of the estimation of the tag population size, because it results in the failure of communication between the reader and tags. In this paper, we propose a coloring graph-based estimation scheme (CGE), which is the first estimation framework designed for multireader and multicategory RFID systems to determine the distribution of tags in different categories. CGE allows for the use of any estimation protocol to determine the number of tags, prevents R2Rc, and results in higher time efficiency and less power-consumption than the classic scheduling method DCS.

## 1. Introduction

Radio Frequency Identification (RFID) technology has been widely deployed to monitor objects in the supply chain. Due to a decrease in the price of tags, RFID is increasingly used in many applications, such as real-time inventory control [1–5] and product tracking [6–9]. A simple RFID system consists of a reader, antennas, and a number of tags. The reader sends the continuous waves to tags, which include operation codes and specifying PHY/MAC parameters. When the tag is activated, it will backscatter a message or keep silent according to the reader's command.

Tag estimation is a fundamental functionality for RFID-enabled warehouse management and inventory control. Most tag estimation protocols estimate the population of a single category of tags based on the slot status observed by a single reader. For example, FNEB [1] uses the number of consecutive empty slots to estimate the number of tags. ES [3] exploits the number of singleton slots occupied in a time frame to estimate the tag cardinality. SEM [10] exploits the

Manchester coding mechanism to simultaneously estimate the tag population of multiple categories.

However, the current large-scale RFID systems usually deployed with multiple readers for tracking a large number of tags from multiple categories. Multicategory tag estimation is particularly challenging in multireader RFID systems because of the reader-to-reader collision (R2Rc) and the multicoverage tags. The R2Rc underestimates the tag populations due to the missing reading caused by communication failure, while a multicoverage tag maybe counted multiple times with distinctive readers. Although some reader-scheduling schemes [11–15] were proposed to resolve the issues of R2Rc and multicoverage tags, their time efficiency is far from optimal because none of them makes use of the tag category distribution characteristics. The tag categories are distributed sequentially, and one category tags are centrally piled up together. Hence, they are of low time efficiency because they have to perform the multicategory RFID estimation category by category [4]. It is noted that in large-scale RFID systems, there are another kind of collision types

named tag-to-tag collision. To cope with the tag-to-tag collision, many works [16–20] have been presented. In this paper, we focus on the R2Rc issue.

Therefore, we propose the coloring graph-based estimation (CGE) scheme to reduce the reader-to-reader collisions to improve the time efficiency of tag cardinality estimation schemes in multiple categories and multiple reader RFID systems. Our proposed scheme CGE can discover the interferences among categories, and the categories without interference can be simultaneously estimated. Hence, the time efficiency of our scheme should be much better than others. The scheme uses the distribution of the categories to activate the related readers and eliminates the overlap between the regions of different categories. So, the categories there no overlap between them could estimate the populations simultaneously. Hence, the time efficiency and power consumption of our scheme should be much better than others. The simulation results show that the proposed CGE results in higher time efficiency and less power consumption than the classic scheduling method distributed color selection (DCS) and variable-maximum distributed color selection (VDCS) [11].

## 2. The Proposed CGE Protocol

**2.1. System Model.** A large-scale RFID system commonly has multiple readers and multicategory tags. The tags communicate with the reader by one-hop transmission. Each tag has an identification ID associated with an object. The ID has two parts: the category ID that indicates the type or manufacturer of the associated object and the member ID that represents the sequence number in the category [5]. The tags can be categorized by the system manager according to the manufacturers or brands.

This paper studies the multicategory tag estimation problem in the multireader system. The system consists of a number of RFID tags with  $\lambda$  categories, which are denoted as  $C_1, C_2, \dots, C_\lambda$ , and  $\mu$  carefully deployed readers, which are denoted as  $R_1, R_2, \dots, R_\mu$ . The number of tags in the distinct categories are denoted by  $n_1, n_2, \dots, n_\lambda$ , respectively. For each category  $C_i, i \in [1, \lambda]$ , we set a required reliability  $\alpha \in (0, 1]$ , a confidence interval  $\beta \in [0, 1)$ , and  $P\{|\hat{n}_j - n_j| \leq \beta\} \geq \alpha, j \in [1, \lambda]$ , where  $\hat{n}_j$  is the estimation of  $n_j$ . To comply with the C1G2 standard [21], we assume that the system uses the framed slotted Aloha as the MAC layer protocol for resolving tag collisions.

**2.2. Detailed Design of CGE.** The proposed CGE protocol consists of four stages: category ID collection, coloring of the categorized region, coloring of the categorized readers, and estimation of the population size. These processes provide information on the distribution of the categorized tags, eliminate the overlaps between different category regions, remove the overlaps between readers, and estimate the population size of tags for each category in the monitoring area.

- (1) Category ID collection: In a large inventory scenario with a variety of goods (e.g., supermarket), the category IDs of goods are known to reader, but the reset-

```

Input: Number of monitoring units:  $m$ ; Number of colors:  $M \leq m$ ; The relationship matrix between units:  $X$ ; Initialize the color of regions:  $Y [0: m-1] = -1$ ;  $flag = 0$ ;
Output: One scheme of graph coloring  $Y$ ;
Function UNCOLLISION(int  $k$ )
  for each  $i \in [0, k-1]$  do
    if  $X[k][i] == 1 \wedge Y[k] == Y[i]$  then
      return false;
    else
      return true;
function GRAPHCOLOR(int  $t$ )
   $p = 0$ ;
  if  $t \geq m$  then
     $flag = 1$ ;
  while  $!flag \wedge p < M$  do
     $Y[t] = p$ ;
     $p = p + 1$ ;
    if UNCOLLISION( $t$ ) then
      GRAPHCOLOR( $t+1$ );
  if  $t == 0$  then
    return null;
  GRAPHCOLOR( $0$ );

```

ALGORITHM 1: The coloring region division.

tlement region of each category of goods is unknown to the reader. To make the tag categories free from reader collisions, we need to obtain the spatial distribution region of each category tags. So, one tag at least for one category needs to be single out in an efficient method, and the tag ID should be extracted to represent the distribution of this category. Therefore, in this stage, we leverage the TPS protocol [22] to rapidly collect category information within the communication range of each reader. The TPS need to be executed multiple rounds to collect the category IDs completely, and the process is terminated when there are no more responses in the frame

During the category ID collection stage, we must confirm that there are no overlaps between readers to ensure that the category ID can be collected completely. Since we know the locations of the  $m$  readers, we can obtain a  $m \times m$  matrix  $X$  that represents the collision relationship between the readers. If  $X_{ij} = 0$ , the readers  $R_i$  and  $R_j$  have no reader collision and can work simultaneously for collecting category IDs; if  $X_{ij} = 1$ , the readers  $R_i$  and  $R_j$  are neighboring readers with overlap regions, and the readers have to operate at different times to avoid a conflict. Given the collision matrix  $X$  of readers, the CGE splits the readers into several groups using Algorithm 1. Algorithm 1 is a greedy algorithm. This moment, the monitoring unit is the communication region of one reader, “ $M$ ” is a variable, and we need to change the value of “ $M$ ” to obtain a coloring scheme. After reader coloring, the reader nodes with the same color have no overlap region and can be operated simultaneously to collect category IDs. As illustrated in Figure 1(a), there are six readers in the overlap regions. The colored node represents the reader in

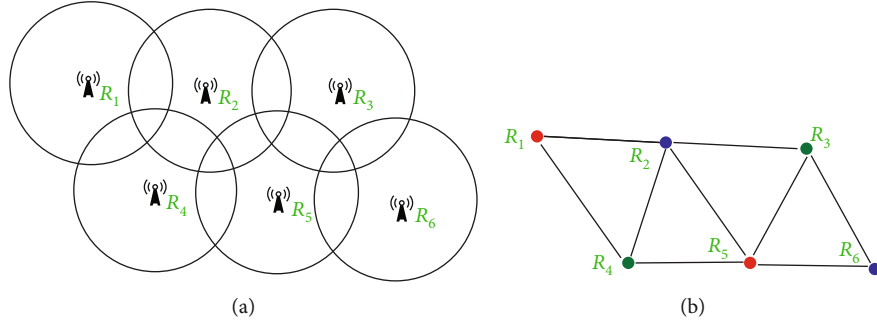


FIGURE 1: An example of reader deployment. (a) Displays a typical RFID system with multireader. (b) A effective batched method in collecting ID phase.

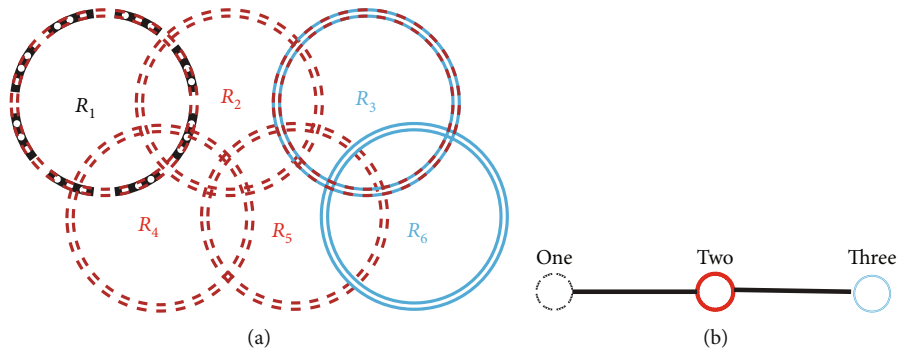


FIGURE 2: Categorized region distribution and the crossing with each other. (a) Categories distribution. (b) The crossing situation of categories distribution regions.

the RFID system, and the line denotes the relationship between the readers. The relationship between the readers is transformed into  $8 \times 8$  binary matrix, where 0 and 1 represent no connection and a connection, respectively. As shown in Figure 1(b), we divide these readers into three colored batches.  $R_1$  and  $R_5$  will collect the category IDs simultaneously in batch one,  $R_2$  and  $R_6$  will be grouped into batch two, and  $R_3$  and  $R_4$  will in batch three.

- (2) Coloring of the categorized region: Each category of tags is monitored by several readers, and we name the monitoring area of these readers categorized region. Due to the population size of each category that is different with each other, the parameters of readers in different categorized region are different, and so we must execute the estimation scheme according to the categorized region

After the first stage, each reader knows the categories within its monitoring region. Subsequently, the distribution region of each category and the relationship between the category nodes are also learned as shown in Figure 2. The category one is monitored by  $R_1$ , the category two is monitored by  $R_1, R_2, R_3, R_4,$  and  $R_5$ , and the category three is monitored by  $R_3$  and  $R_6$ . Again, the server runs Algorithm 1 to find the groups of categories that do not have category-to-category conflicts. The input matrix  $X$  in this stage is the relationship between the cat-

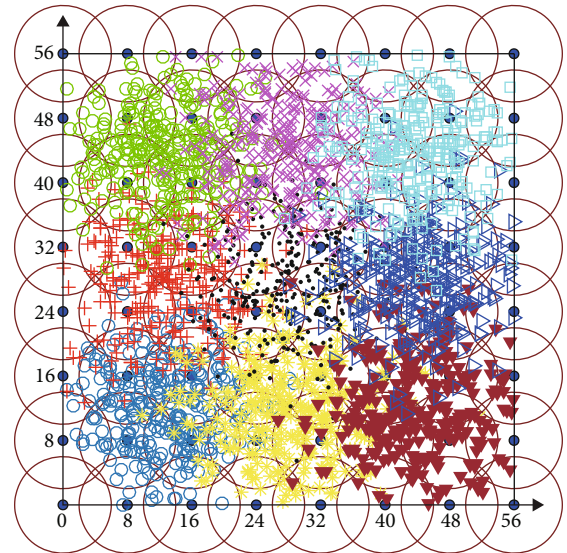


FIGURE 3: The composition of a large-scale RFID system.

egorized nodes,  $M$  is the number of colors, and the  $m$  is the number of categorized regions. The result is that category one and category three are in the same group, and category two is in another batch. These two groups will be scanned sequentially, and categories one and three will be scanned simultaneously.

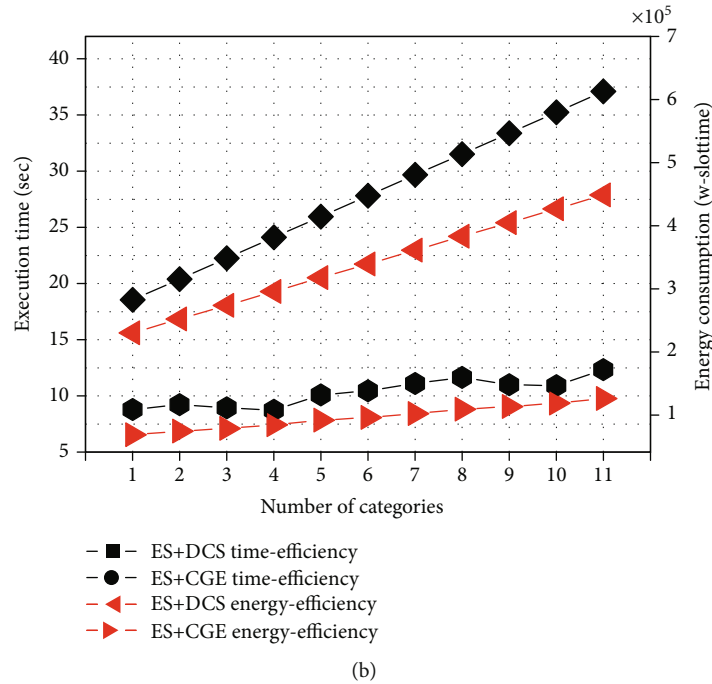
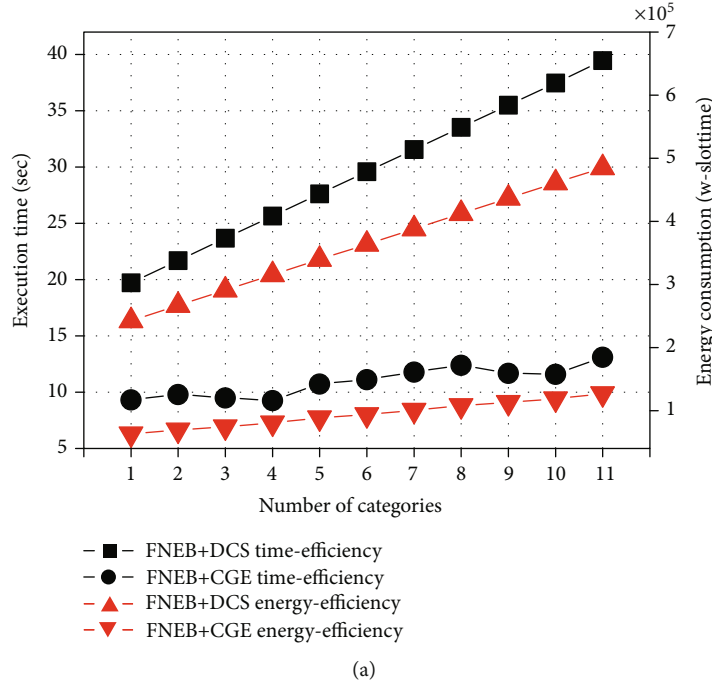


FIGURE 4: The comparisons between FNEB and ES based on our ECG and classic method DCS with the number of categories increasing.

(3) Coloring of the categorized readers: We assign names to the readers that monitor the tags in the same category; these are referred to as categorized readers. Although the work sequence of categories was determined in the previous phase, a single category of tags may be covered by multiple readers in the overlap regions. The overlap between readers causes reader-to-reader collisions (R2Rc), results in identification failure, and decreases the accuracy

of tag estimation. To prevent R2Rc, we have to rearrange the order of operation of the readers. We use the Algorithm 1 to divide the readers into several groups in one categorized region. The input matrix  $X$  in this stage is the relationship between the categorized readers in one categorized region. Within each group, there are no overlaps between readers. Here, the  $M$  is the number of colors, and  $m$  is number of readers in each categorized



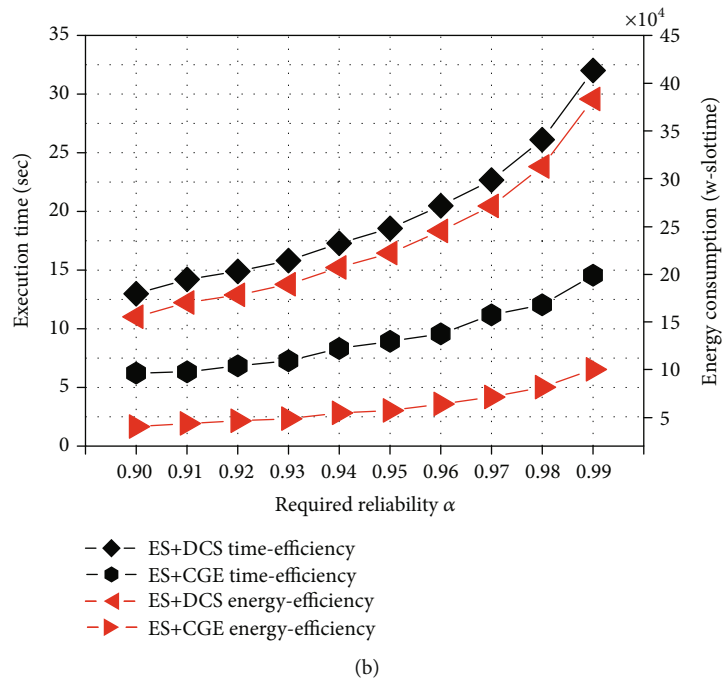
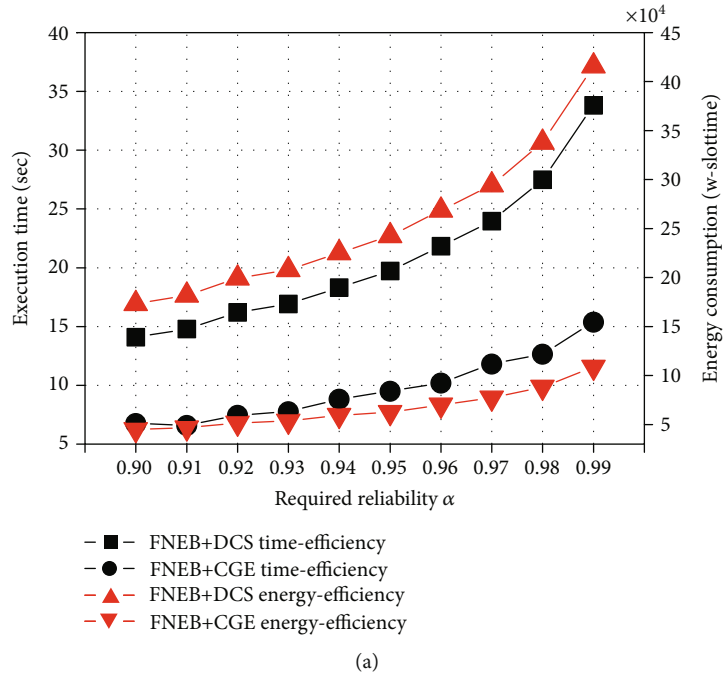


FIGURE 5: The comparisons between FNEB and ES based on our ECG and classic method DCS with  $\alpha$  increasing.

distribution region. As illustrated in Figure 2(a), the readers in category two region are split into three groups:  $B_1$ ,  $B_2$ , and  $B_3$ . The categorized readers  $R_1$  and  $R_5$  in  $B_1$  will work simultaneously,  $R_3$  and  $R_4$  in  $B_2$  and  $R_2$  in  $B_3$ , and the working scheme of the categorized readers in  $B_2$  and  $B_3$  is the same as in  $B_1$ .  $B_1$ ,  $B_2$ , and  $B_3$  will work in sequence to avoid the R2Rc. The readers in categories one and three adopt the same method to split the readers. After that, every reader in the RFID

system is assigned the work sequence. For example, the six readers in Figure 2(a), one scheme is  $R_1$ , and  $R_3$  work simultaneously in first,  $R_6$  will work after  $R_1$  and  $R_3$  finish, and  $R_6$  is done representing the estimation results of categories one and three that are obtained. The work sequence of readers in category two that has been introduced above

- (4) Estimation of the population size: After the former three stages, all the readers have been arranged to

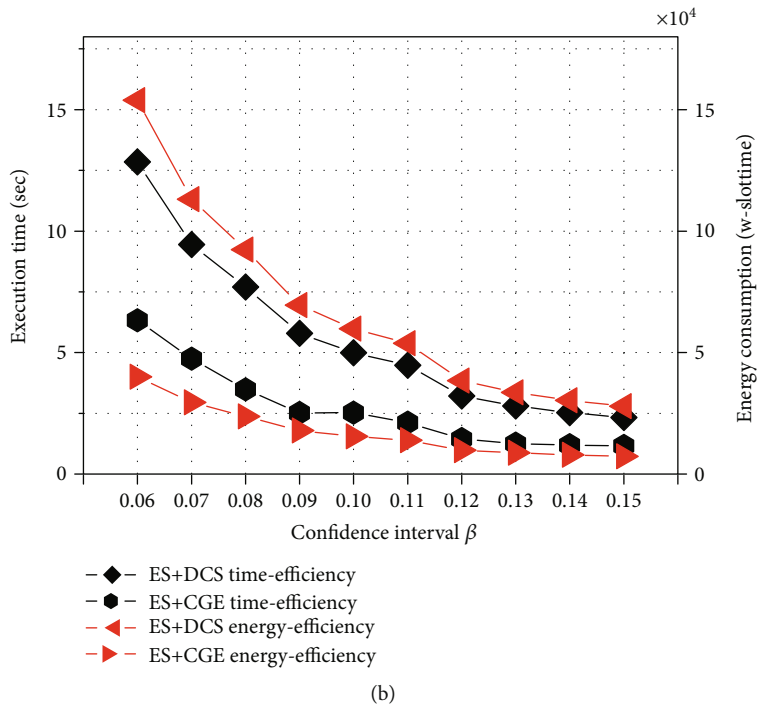
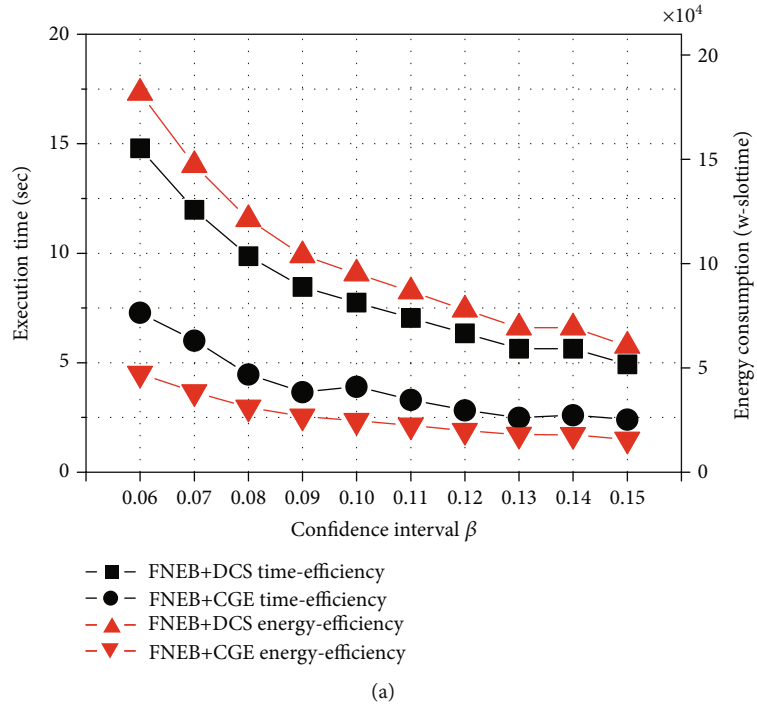


FIGURE 6: The comparisons between FNEB and ES based on our ECG and classic method DCS with  $\beta$  increasing.

run in sequence for avoiding R2Rc. Since the tags in the overlapping zone covered by multiple readers will be counted more than once, the tag cardinality will be overestimated. To overcome this issue, we let all readers in one categorized region broadcast the same parameters. Since the tags use the hash function to select time slot to respond, tag's responses will appear in the slots of related frames with same index. Each

reader generates a binary frame vector, in which bit 1 indicates a busy slot and bit 0 indicates an empty slot. Then, we use OR operation to aggregate the frame vectors received from readers. Owing to the characteristics of the OR operation, multiple responses of a tag will only create one bit 1 in the aggregated frame vector. Thus, using the aggregated frame vector, the tag cardinality will not be

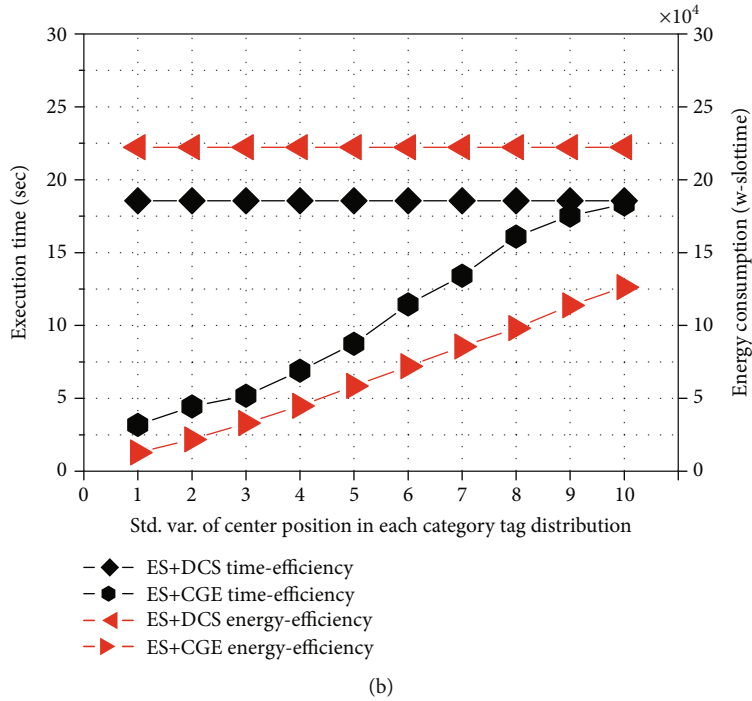
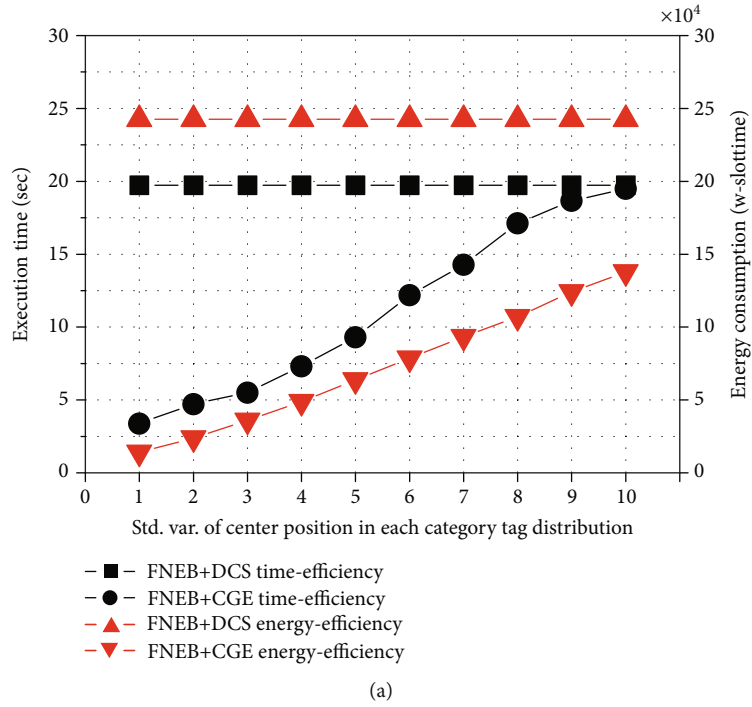


FIGURE 7: The comparisons between FNEB and ES based on our ECG and classic method DCS with  $\sigma$  increasing.

overestimated. In this stage, the previous estimation protocols FNEB [3], ES [1], etc. can be used in the ECG framework to estimate the population size of the tags. Unlike the exact tag identification schemes that need to resolve the tag-to-tag collisions, some probabilistic tag estimation schemes can benefit from the tag collisions, e.g., ART [23] and ZOE [24]. Hence, we do not optimize the frame size to reduce the tag-to-tag collisions. To ensure the convergence

of the results, we run the experiments independently for multiple rounds (i.e., 1000) and then calculate the average value [26–30]

### 3. Performance Evaluation

In this section, we demonstrate the performance of the proposed CGE. In the simulation, we implement the estimation protocols FNEB and ES based on our CGE scheduling

scheme and the classic reader scheduling method DCS and VDCS under the same parameters [11]. The precondition of the comparisons is that the locations of readers are known to us. For a fair comparison, each protocol uses the same wireless communication setting as follows. In term of time efficiency, the tag-to-reader transmission rate is 53Kb/s; otherwise, it is 26.5Kb/s. That is, it takes  $18.8\mu\text{s}$  to transmit one bit from a tag to the reader, and  $37.7\mu\text{s}$  to transmit one bit from a reader to a tag. Between two consecutive slots, there is a waiting time  $\tau_w = 302\mu\text{s}$  [8]. In terms of energy efficiency, the power of one query from reader is 825 mW, and the power of one bit response from reader is 125 mW [25]. In the experiment, we set the required reliability  $\alpha$  that is 0.95, the confidence interval  $\beta$  is 0.05, the number of categories  $\lambda$  is 9, the number of tags in each category  $n_i$  is 1000, and the standard variance of the center position in each categorized tag distribution  $\sigma$  is 3. The communication radius of the reader is 6 m. In order to promote the understanding of readers, we show an example to illustrate the deployment of readers and the distribution of categories in Figure 3. The width and length of the monitoring area are 56 m and 56 m, respectively, and the communication radius of the reader is 6 m.

**3.1. The Time Efficiency of Estimation.** From Figures 4–7, we discover several observations as follows. First, the performance of the proposed CGE framework is better than that of the DCS and VDCS in terms of time efficiency. For example, when the data of required reliability  $\alpha$  is 0.99, the time cost of the FNEB+DCS and ES+DCS is 33.44 s and 28.81 s, respectively. And the time cost of the FENB+VDCS and ES+VDCS is 37.23 s and 26.24 s, respectively. However, the time cost of the FNEB+CGE and ES+CGE is 13.53 and 12.80, respectively. It means that the CGE framework makes the FNEB and ES acquire  $2.25\times$  faster than DCS and  $2.05\times$  faster than VDCS at least. Second, the estimation time of the FNEB and ES increases as  $\lambda$ ,  $\alpha$ , increases, because the higher  $\lambda$  and  $\alpha$  mean the stricter demand for achieving the required performance. In contrast, the estimation times of the FNEB and ES decrease as  $\beta$  increases. Third, the change speed of the FNEB+CGE and ES+CGE is lower than that of the FNEB+DCS and ES+DCS. Finally, the performance of the CGE depends on the distribution of the tag categories. The commodities, which tags are attached to, are more concentrated, and the time efficiency is higher which we discovered in Figure 7.

**3.2. The Energy Efficiency of Estimation.** In terms of energy efficiency, we discover three observations from Figures 4–7 as follows. First, the power consumption of the proposed CGE is lower than that of the classic scheduling method DCS. The power consumption of the ECG framework in the FNEB and ES methods is 72% lower than that of the DCS method, when the required reliability  $\alpha$  of the RFID system is 0.99. Second, the energy cost of the FNEB and ES increases as  $\lambda$  and  $\alpha$  increase and decrease as  $\beta$  increases. Finally, the power consumption of the DCS is not affected by the category distribution; however, the CGE takes advantage of the distribution of the categories, and multiple catego-

ries can be estimated simultaneously in one group. The simulation results show that the proposed CGE framework for estimating the tag population size results in higher time efficiency and less power consumption than the DCS method.

## 4. Conclusion

This paper proposed the CGE protocol for tag estimation in multireader and multicategory RFID systems. CGE discovers the interferences among categories and eliminates the reader-to-read collision by coloring the parallelizable categorized regions with the same color in two stages. CGE promotes the time efficiency of existing estimation algorithm because more regions can be parallelized for tag estimation in the multireader RFID system. And the simulation results illustrate the superiority of CGE under various parameters setting compared with other scheduling protocol.

## Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The author(s) declare(s) that they have no conflicts of interest.

## References

- [1] L. Xie, H. Han, Q. Li, J. Wu, and S. Lu, "Efficiently collecting histograms over RFID tags," in *IEEE INFOCOM 2014 - IEEE Conference on Computer Communications*, Toronto, ON, Canada, 2014.
- [2] J. Su, Z. Sheng, V. C. M. Leung, and Y. Chen, "Energy efficient tag identification algorithms for RFID: survey, motivation and new design," *IEEE Wireless Communications*, vol. 26, no. 3, pp. 118–124, 2019.
- [3] H. Han, B. Sheng, C. C. Tan, Q. Li, W. Mao, and S. Lu, "Counting RFID tags efficiently and anonymously," in *Proceedings IEEE INFOCOM*, San Diego, CA, USA, 2010.
- [4] S. Zhang, X. Liu, J. Wang, and J. Cao, "Tag size profiling in multiple reader RFID systems," in *IEEE INFOCOM 2017 - IEEE Conference on Computer Communications*, Atlanta, GA, USA, 2017.
- [5] X. Liu, B. Xiao, K. Li et al., "RFID Cardinality Estimation with Blocker Tags," in *2015 IEEE Conference on Computer Communications (INFOCOM)*, Hongkong, China, 2015.
- [6] J. Su, Z. Sheng, A. Liu, Z. Fu, and Y. Chen, "A time and energy saving based frame adjustment strategy (TES-FAS) tag identification algorithm for UHF RFID systems," *IEEE Transactions on Wireless Communications*, vol. 19, no. 5, pp. 2974–2986, 2020.
- [7] J. Su, Y. Chen, Z. Sheng, Z. Huang, and A. X. Liu, "From M-ary query to bit query: a new strategy for efficient large-scale RFID identification," *IEEE Transactions on Communications*, vol. 68, no. 4, pp. 2381–2393, 2020.
- [8] X. Liu, K. Guo, Z. Liu, X. Zhou, H. Qi, and W. Xue, "Fast and accurate missing tag detection for multi-category RFID

- systems,” in *2018 IEEE International Conference on Smart Internet of Things (SmartIoT)*, Xi'an, China, 2018.
- [9] X. Liu, S. Chen, J. Liu et al., “Fast and accurate detection of unknown tags for RFID systems – hash collisions are desirable,” *IEEE/ACM Transactions on Networking*, vol. 28, no. 1, pp. 126–139, 2020.
- [10] X. Liu, K. Li, A. X. Liu et al., “Multi-category RFID estimation,” *IEEE/ACM Transactions on Networking*, vol. 25, no. 1, pp. 264–277, 2017.
- [11] J. Waldrop, D. W. Engels, and S. E. Sarma, “Colorwave: an anticollision algorithm for the reader collision problem,” in *IEEE International Conference on Communications, 2003. ICC '03*, vol. 2, pp. 1206–1210, Anchorage, AK, USA, 2003.
- [12] F. Gandino, R. Ferrero, B. Montrucchio, and M. Rebaudengo, “Probabilistic DCS: an RFID reader-to-reader anti-collision protocol,” *Journal of Network and Computer Applications*, vol. 34, no. 3, pp. 821–832, 2011.
- [13] A. A. Mbacke, N. Mitton, and H. Rivano, “A survey of RFID readers anticollision protocols,” *Journal of Network and Computer Applications*, vol. 2, no. 1, pp. 38–48, 2018.
- [14] I. Amadou, N. Mitton, and S. E. Sarma, “HAMAC: high adaptive MAC protocol for dense RFID reader-to-reader networks,” in *International Conference on Ad Hoc Networks (AdHocNets)*, San Remo, Italy, 2015.
- [15] J. Su, Z. Sheng, A. X. Liu, Y. Han, and Y. Chen, “A group-based binary splitting algorithm for UHF RFID anti-collision systems,” *IEEE Transactions on Communications*, vol. 68, no. 2, pp. 998–1012, 2020.
- [16] J. Su, A. X. Liu, Z. Sheng, and Y. Chen, “A partitioning approach to RFID identification,” *IEEE/ACM Transactions on Networking*, vol. 28, no. 5, pp. 2160–2173, 2020.
- [17] J. Su, Z. Sheng, A. X. Liu, Y. Han, and Y. Chen, “Capture-aware identification of mobile RFID tags with unreliable channels,” *IEEE Transactions on Mobile Computing*, p. 1, 2020.
- [18] J. Su, R. Xu, S. Yu, B. Wang, and J. Wang, “Idle slots skipped mechanism based tag identification algorithm with enhanced collision detection,” *KSII Transactions on Internet and Information Systems*, vol. 14, no. 5, 2020.
- [19] Z. Kuang, L. Li, J. Gao, L. Zhao, and A. Liu, “Partial offloading scheduling and power allocation for mobile edge computing systems,” *IEEE Internet of Things Journal*, vol. 6, no. 4, pp. 6774–6785, 2019.
- [20] Z. Kuang, L. Zhang, L. Zhao, and L. Zhao, “Energy- and spectral-efficiency tradeoff with  $\alpha$ -Fairness in energy harvesting D2D communication,” *IEEE Transactions on Vehicular Technology*, vol. 69, no. 9, pp. 9972–9983, 2020.
- [21] EPCglobal, “EPC radio-frequency identity protocols class-1 generation-2 UHF RFID protocol for communications at 860 mhz-960mhz version 1.2.0,” Epcglobal, Brussels, Belgium, 2015.
- [22] J. Liu, S. Chen, B. Xiao, Y. Wang, and L. Chen, “Category information collection in RFID systems,” in *IEEE 37th International Conference on Distributed Computing Systems (ICDCS)*, Atlanta, GA, USA, 2017.
- [23] S. Muhammad and A. X. Liu, “Every Bit Counts: Fast and Scalable RFID Estimation,” in *Proceedings of the 18th annual international conference on Mobile computing and networking*, Istanbul, Turkey, 2012.
- [24] Y. Zheng and M. Li, “ZOE: fast cardinality estimation for large-scale RFID systems,” in *2013 Proceedings IEEE INFOCOM*, Turin, Italy, 2013.
- [25] V. Namboodiri and Lixin Gao, “Energy-aware tag Anticollision protocols for RFID systems,” *IEEE Transactions on Mobile Computing*, vol. 9, no. 1, pp. 44–59, 2010.
- [26] Z. Zhou, Q. M. Wu, Y. Yang, and X. Sun, “Region-level visual consistency verification for large-scale partial-duplicate image search,” *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 16, no. 2, pp. 1–25, 2020.
- [27] Z. Zhou, Y. Mu, and Q. M. Wu, “Coverless image steganography using partial-duplicate image retrieval,” *Soft Computing*, vol. 23, no. 13, pp. 4927–4938, 2019.
- [28] X. Liu, J. Zhang, S. Jiang et al., “Accurate localization of tagged objects using mobile RFID-augmented robots,” *IEEE Transactions on Mobile Computing*, vol. 20, no. 4, pp. 1273–1284, 2021.
- [29] J. Su, R. Xu, S. Yu, B. Wang, and J. Wang, “Redundant rule detection for software-defined networking,” *KSII Transactions on Internet and Information Systems*, vol. 14, no. 6, 2020.
- [30] J. Su, Z. Sheng, A. X. Liu, Z. Fu, and C. Huang, “An efficient missing tag identification approach in RFID collisions,” *IEEE Transactions on Mobile Computing*, p. 1, 2021.

## Research Article

# Intelligent Recognition and Teaching of English Fuzzy Texts Based on Fuzzy Computing and Big Data

Ling Liu <sup>1</sup> and Sang-Bing Tsai <sup>2</sup>

<sup>1</sup>*School of Foreign Languages, Gannan University, Ganzhou, Jiangxi 341000, China*

<sup>2</sup>*Regional Green Economy Development Research Center, School of Business, Wuyi University, China*

Correspondence should be addressed to Ling Liu; [benbenbianbian1615@163.com](mailto:benbenbianbian1615@163.com) and Sang-Bing Tsai; [sangbing@hotmail.com](mailto:sangbing@hotmail.com)

Received 2 June 2021; Revised 20 June 2021; Accepted 1 July 2021; Published 12 July 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Ling Liu and Sang-Bing Tsai. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this paper, we conduct in-depth research and analysis on the intelligent recognition and teaching of English fuzzy text through parallel projection and region expansion. Multisense Soft Cluster Vector (MSCVec), a multisense word vector model based on nonnegative matrix decomposition and sparse soft clustering, is constructed. The MSCVec model is a monolingual word vector model, which uses nonnegative matrix decomposition of positive point mutual information between words and contexts to extract low-rank expressions of mixed semantics of multisense words and then uses sparse. It uses the nonnegative matrix decomposition of the positive pointwise mutual information between words and contexts to extract the low-rank expressions of the mixed semantics of the polysemous words and then uses the sparse soft clustering algorithm to partition the multiple word senses of the polysemous words and also obtains the global sense of the polysemous word affiliation distribution; the specific polysemous word cluster classes are determined based on the negative mean log-likelihood of the global affiliation between the contextual semantics and the polysemous words, and finally, the polysemous word vectors are learned using the Fast text model under the extended dictionary word set. The advantage of the MSCVec model is that it is an unsupervised learning process without any knowledge base, and the substring representation in the model ensures the generation of unregistered word vectors; in addition, the global affiliation of the MSCVec model can also expect polysemantic word vectors to single word vectors. Compared with the traditional static word vectors, MSCVec shows excellent results in both word similarity and downstream text classification task experiments. The two sets of features are then fused and extended into new semantic features, and similarity classification experiments and stack generalization experiments are designed for comparison. In the cross-lingual sentence-level similarity detection task, SCLVec cross-lingual word vector lexical-level features outperform MSCVec multisense word vector features as the input embedding layer; deep semantic sentence-level features trained by twin recurrent neural networks outperform the semantic features of twin convolutional neural networks; extensions of traditional statistical features can effectively improve cross-lingual similarity detection performance, especially cross-lingual topic model (BL-LDA); the stack generalization integration approach maximizes the error rate of the underlying classifier and improves the detection accuracy.

## 1. Introduction

The important means of education informatization is to apply information technology and network technology to education to realize the mode of “Internet + education” [1]. Education informatization covers various aspects such as education management, education process, and education resources. In terms of educational resources, in addition to physical/digital forms such as paper and digital textbooks,

tutorial materials, teaching guides, and practice problems, there is a large amount of data, information, and resources in the network that can provide support for education informatization. In web-based education, learning resources can be shared and students can have a more generalized learning space. In the new informatization education platform, the learning place has changed from the fixed classroom and fixed class time to anytime and anywhere teaching, and the learning means has changed from traditional blackboard

and paper books to multimedia video and audio and electronic courseware, etc. The education form is getting richer and richer, and network education starts to play an increasingly important role. According to different concepts and objects, online education can be divided into online education, distance education, adaptive education, virtual classroom, and other forms.

In addition to the above roles of text similarity detection, its direct application has considerable practical value [2]. For example, it has a significant role in protecting the intellectual property rights of electronic texts and combating illegal copying and plagiarism of academic results. Since its launch, the “Academic Misconduct Detection System” of China Knowledge Network has blocked the publication of articles with high repetition rate at the source of academic results; the CrossCheck antiplagiarism literature detection system developed by the International Publishing Links Association has minimized the textual repetition rate of English publications [3]. The above two systems are both text similarity-based detection systems, which shows that the technology has important application value and social significance for combating paper forgery, correcting academic culture, and promoting independent innovation. Unfortunately, most of the successful text similarity detection systems are monolingual, and cross-lingual similarity detection systems have been tried, but the detection effect is poor [4]. The frequency-based word packet model constructs text feature engineering only by the frequency or number of words appearing in the document which mainly leads to the advantage of words with higher frequency or number in the text data in the text feature extraction process.

This thesis investigates data mining methods and application research for personalized learning, providing data-driven solutions. The research objects are learners, learning resources, and personalized learning mechanisms. Learners are broadly defined as students who are engaged in learning activities (including online students and offline students whose academic activities are recorded). Learning resources are the practice questions (including daily homework questions, test questions, and general classroom exercises) that have been recorded by various products and systems and are available to students for learning. Learning strategies are broadly defined as those that can support and ensure learners perform personalized learning recommendation strategies, focusing on, for example, online practice question recommendation algorithms. Although the relevant research on relation extraction is relatively adequate, based on the above analysis and introduction, the extraction of semantic relations from short English texts is highly differentiated from the classical relation extraction task. Studying this topic is of great importance both at the algorithmic level and at the application level.

## 2. Current Status of Research

Calzada et al. proposed the LexVec model, which uses a sampling-based approach to decompose the PPMI matrix and reconstructs the loss function to penalize excessive word-frequency cooccurrence errors [5]. PPMI-SVD, on

the other hand, uses the truncated singular value decomposition (SVD) technique to achieve dimensionality reduction, aimed at finding the maximum features to describe the semantics of words and the retained dimensional features to approximate the PPMI matrix composed of word-context information [6]. Charitopoulos et al. point out that the SVD rank reduction algorithm does not guarantee the nonnegativity of the matrix decomposition, so nonnegative matrix factorization (NMF) becomes a sensible choice for decomposing the PPMI, which guarantees the nonnegativity of the dimensional approximate reduction of the semantic relationship of the word-context information and is more consistent with the semantic relationship hypothesis [7]. Raveh et al. demonstrated that matrix factorization methods are consistent with the word vectors of neural network language models in terms of task performance, such that PPMI-SVD matrix decomposition is equivalent to skip-gram (SGNS) based on negative sampling [8]. The neural network language model has an inherent advantage over factorization methods, which have more flexible super parameter settings [9].

With the development of the Internet, the dramatic increase of all kinds of information, especially textual information, has brought great challenges to the management of knowledge and the mining of relationships among knowledge [10]. There is an urgent need to retrieve the target information more quickly and efficiently and even expect intelligent access to all information related to the target [11]. To meet this demand, knowledge bases are created, in which ontologies are one of the means to manage knowledge more conveniently and to explore the relationships between knowledge [12]. As creating ontologies manually is not very practical in ontology engineering, it is time-consuming to build ontologies manually especially when the amount of data is large. While semiautomatic construction is difficult to reconcile with manual construction and also suffers from subjective problems, many scholars have started to try to develop automatic ontology creation methods [13]. For example, Levi et al. proposed a novel event-based ontology construction method that extracts domain ontologies mainly from unstructured text data [14]. We know that there is rich knowledge in many data, such as text and databases [15]. The automatic ontology construction method automatically extracts knowledge from these data to conform to a specific format (ontology). However, the results are always unsatisfactory because the ontologies generated by the automatic process are not accurate and of low quality, so automatic construction is still a challenge that is not yet perfect.

In contrast, semiautomatic construction is based on a combination of manual and automatic methods, which needs to coordinate the balance between manual and machine, and the semiautomatic approach will encounter the problems existing in both manual and automatic [16]. The automatic ontology construction approach can effectively extract knowledge from a large amount of text to build ontologies based on such knowledge, but the difficulty of automatic construction is that it requires the builder to be knowledgeable about the domain and the developer to have a proficiency in the key technologies involved in the construction [17].

Whether semiautomatic or automatic construction, the key technologies involved mainly include knowledge representation, entity recognition, relationship extraction, graph storage, relationship inference, entity disambiguation, and other technologies. Building an ontology requires systematic integration of these technologies to form a complete set of processes. In traditional semiautomatic construction methods, it is necessary to rely on linguistic tools or artificially construct many features (including lexical and semantic features) and then train a machine learning model for each task (e.g., entity recognition and relation extraction). The disadvantage of this approach is that the applicability of the model is low and the performance of the model depends heavily on the quality of the features. With the development of deep learning, neural network-based methods have been proposed, but these methods still have some shortcomings, such as the extraction of multiple relationships and the problem of overlapping entity labels. Given the shortcomings in the key techniques of the above ontology construction methods, this paper conducts an in-depth study and proposes a solution with better performance.

### 3. Analysis of Parallel Projection and Region-Expanding English Fuzzy Text Intelligent Recognition Teaching

*3.1. English Fuzzy Text Intelligent Recognition Algorithm Design.* Text mining is a mathematical modeling approach to characterize the important relationship between internal rules and the semantic level of text by using mathematical statistics as a general idea. However, as nonnumerical information, text cannot be directly used as input data for computer modeling languages to obtain results. Therefore, converting text-based data into computer-readable numerical data becomes a prerequisite and foundation for text mining. On the one hand, the bag-of-words model based on word frequency constructs text feature engineering only by the frequency or frequency of words appearing in the document, which leads to the fact that word items with higher frequency or frequency in text data will have a natural advantage in the text feature extraction process [18]. When extracting features from documents, using frequency alone as the measure of word item importance can result in the absence of key information in the text. In text data, certain words or phrases may occur infrequently, but still contain key information of the document. In general, the frequency of deactivated words in the text is much higher than that of keywords that can reflect the theme of the document, so removing deactivated words will significantly improve the accuracy of feature extraction; on the other hand, when the volume of text data is large, this high-dimensional sparse matrix representation will bring a great difficulty to the model operation.

The training data consists of historical observations of the explanatory variables and the corresponding response variables. The model can predict the values of the response variables corresponding to explanatory variables that are not in the training data. The goal of the regression problem is to predict the continuous values of the response variables.

Several linear regression models will be learned, and the training data, modeling and learning algorithms, and evaluation of the effectiveness of each method will be presented later.

The word frequency-inverse document frequency model, abbreviated as TF-IDF model, which combines the word frequency of words in text data with their inverse document frequency, started as a metric to show the ranking function of search engine user query results and has now become one of the important methods for information retrieval and text feature extraction. The mathematical representation of the model is shown in

$$\text{tf} + \text{idf} = \text{tf} \times \text{idf}, \quad (1)$$

where  $\text{tf}$  denotes the word frequency of a text unit,  $\text{idf}$  denotes its inverse document frequency, so the  $\text{tf-idf}$  value of the text unit is the product of the two. First, the word frequency  $\text{tf}$  is the frequency value of a word in a particular document; the frequency value has a variety of different ways to represent, either its absolute frequency value or frequency value or can be a binary variable, that is, the word in the document is expressed as 1, and vice versa is 0. Suppose there is a word set  $w$  and document set  $D$ , as shown in equation (2), then the frequency value of the word  $w$ , in document  $D$ , which is the  $\text{tf}$  value is shown in equation (3).

$$w = \{w_1, w_2, w_3, \dots, w_k\}, \quad (2)$$

$$D = \{D_1, D_2, D_3, \dots, D_k\}, \quad (3)$$

$$\text{tf}(w_i, D_j) = N_{w_i, D_j} \sum_{n=1}^K N_{w_n, D_j} N, \quad (4)$$

where  $N_{w_i, D_j}$  denotes the frequency of text units  $w$ , in document  $D$ , and the denominator denotes the total frequency of each word in the dictionary in document  $D_j$ . Second, the inverse document frequency  $\text{idf}$  is the inverse of the document frequency of each text unit, obtained by dividing the total number of documents by the number of documents containing the specified text unit and then taking the logarithm of the resulting quotient. The  $\text{idf}$  value is calculated as shown in

$$\text{idf}(w_i) = \ln \frac{|D|}{|\{D_i : D_j \in w_j\} - 1|}. \quad (5)$$

The TF-IDF model based on the bag-of-words model achieves feature extraction of documents based on statistical word frequency combined with inverse document frequency. However, text data belongs to sequence-type information, and the position relationship of text units has a very important influence on semantics, so it is impossible to understand the meaning of text content accurately by relying solely on frequency-based features. In particular, in natural language processing tasks, the preceding and following texts are very closely related, so each word, phrase, or sentence cannot be analyzed in isolation [19]. Exploring text feature extraction



models based on sequence information, combining frequency characteristics with text unit location information, has become an important research direction for document feature engineering construction. In this paper, we adopt the document word vector training model based on an artificial neural network and combine it with the TF-IDF model to propose the tf-idf weighted average word vector to represent document features, as shown in Figure 1.

The key point of the fitting process is whether the values of the parameters such as the weight term and bias term in the model satisfy the corresponding conditions. The proposed training algorithm achieves automatic learning of the relevant parameters, and the model can automatically modify the values of the weight terms and bias terms by training on the sample data, thus solving the problem of poor generality of human settings. The algorithm first initializes the values of the parameters to 0 and then modifies them using the perceptron iteration rules, which are shown in

$$w_i = w_i - \eta(t - y)x_i, \quad (6)$$

$$b = b - \eta(t - y). \quad (7)$$

$w_i$  is the weight term corresponding to the input information  $x_i$ ,  $b$  is the bias term,  $t$  is the actual value of the training sample,  $y$  is the output value calculated by the model, and  $\eta$  is the learning rate, which is used to adjust the magnitude of each weight adjustment. In the training algorithm, the model calculates the output value corresponding to each training data in turn and adjusts the weight term accordingly using the above iteration rule. By calculating the sample data in turn and iteratively modifying the weight terms, the training set data are processed iteratively to finally obtain the relevant parameters of this perceptron model, to achieve an accurate fit to the target function.

When used as an independent unit, a perceptron can fit a linear function; when used as a component in an artificial neural network, a perceptron is called a neuron. These neurons are interconnected according to certain rules to form a complete neural network model. Each neuron in the model is computed in the same way as the perceptron. The structure of a classical fully connected neural network is shown in Figure 2.

The neurons in the neural network model are laid out in a layer structure. The leftmost layer in Figure 2 is the input layer with three nodes, which is responsible for receiving the input data; the rightmost layer is the output layer with two nodes, which is responsible for the output of the model; the layer between the input layer and the output layer is called the hidden layer with four nodes, which is responsible for the intermediate transmission of the data. The complexity of our model is relatively low, and the performance of the general configuration can meet the demand; of course, the higher the performance the faster the time. In the fully connected neural network model shown above, all neurons in adjacent layers are connected, i.e., each neuron in layer IV is connected to all neurons in layer N-1, and the output of the neuron in layer N-1 is the input of the neuron in layer IV; meanwhile, there is no connection between neurons in

the same layer, and each neuron is individually numbered, and the operation of this neural network model as the algorithm of this neural network model is described as follows. First, the output value of the nodes in the input layer is the input vector  $X$ . Therefore, the dimension of the input vector should be the same as the number of neurons in the input layer; second, after obtaining the output value of each neuron in the input layer, the output value of each node in the hidden layer can be calculated. Take note of the hidden layer as an example, its calculation method is shown in

$$a = f(W^T \otimes X). \quad (8)$$

After the above steps, the artificial neural network model can obtain the predicted values of the sample data input to the model, thus achieving the purpose of supervised machine learning. In addition to the most basic three-layer fully connected neural network model, many variants such as deep neural networks, convolutional neural networks, and recurrent neural networks have emerged after a long period of research and development. The number of hidden layers is changed from a single layer to multiple layers, and with the increase of the number of hidden layers, the fitting of the algorithm model will show better results, but the difficulty of the model training and the computing time will increase exponentially, so there are high requirements on the computer hardware conditions. In deep learning with the deep neural network as the core, the number of hidden layers should be controlled within 10 layers in general. A sentence contains a lot of information, but it cannot have a corresponding meaning when separated by a word live phrase.

The  $N$ -gram method is based on the hidden Markov chain assumption that the text information at the current position is only relevant to a finite number of text units before that position, without having to take all the previous contents into account [20]. According to such an assumption, a large-scale reduction of the training data can be achieved, thus improving the training efficiency of the model. In the actual training data construction process, not only the content before the specified position will be considered but also the text data after that position, and the window sizes before and after will be kept consistent. In the  $N$ -gram method, the size of the window  $N$  is artificially set and is, therefore, one of the hyperparameters of the model. For general text data, the best empirical value for the window size is 5, i.e., the information of 5 text units before and after the specified position is considered.

The process of constructing the model training data in this paper uses the 5-gram method, in which the word vector of the specified text unit generated using the method in the previous subsection is used as the sample input feature, and the word vector of the text unit whose distance from the unit is within the window size (5) is used as the label of the sample to construct the model training data. In this subsection, to provide a brief explanation of the algorithm principle, only the first sentence of the sample text data is used as an example, and a 1-gram is used for the construction of the training

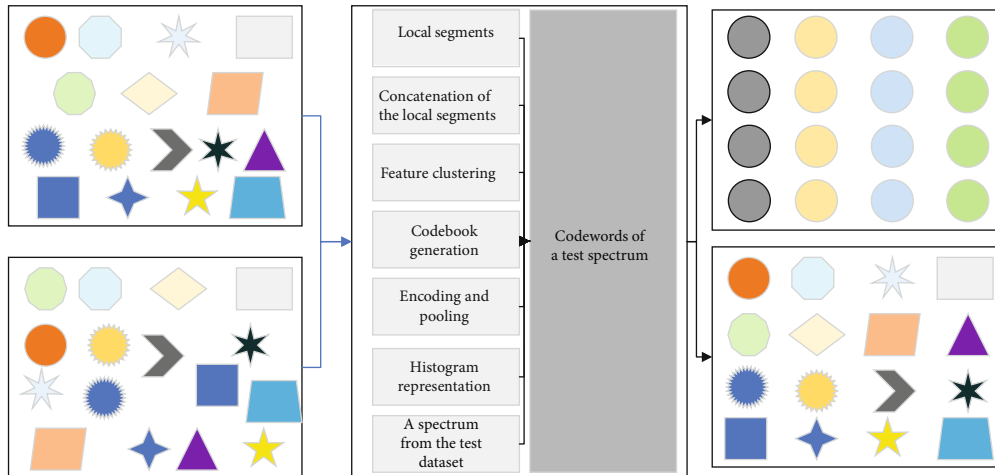


FIGURE 1: The framework of the bag-of-words model.

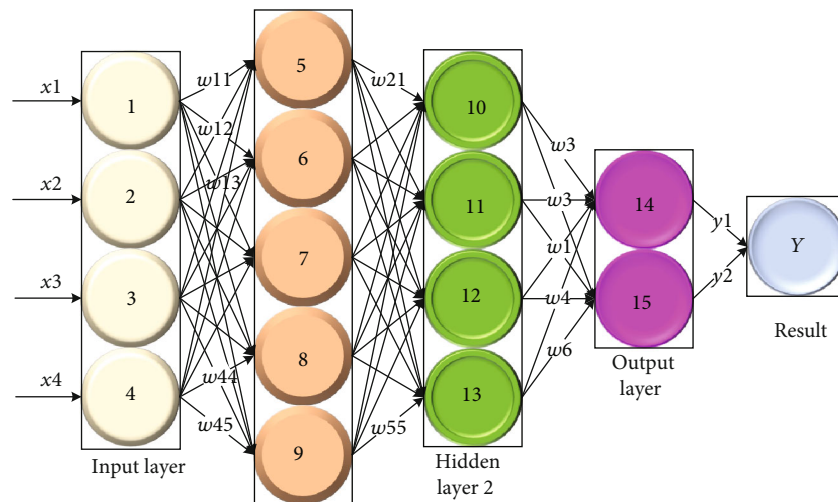


FIGURE 2: Structural diagram of fully connected neural network.

data, and the construction process and the generated results are shown in Figure 3.

The advantages of converting computer-unreadable text data into computer-readable digital information are mainly easier to process and analyze. According to the model of this paper, the purpose of constructing document feature engineering in terms of sentences is achieved. The model combines the principles of artificial neural networks and their algorithms to achieve feature extraction of documents by creating word vectors for independent words. The advantage of the proposed method in this paper is that it not only preserves the frequency information of text units in the document but also combines the location information between different text units, thus enabling the model to show a better fit when dealing with sequence-type text data. The document feature extraction work is the prerequisite and foundation of the higher-order text mining task, and the process converts computer-unreadable text-based data into computer-readable numerical information. The construction of document feature engineering lays the foundation for the key

information extraction based on text distance and the intelligent recognition algorithm in the later paper.

### 3.2. Parallel Projection and Regional Extension ELT Design.

The smart classroom environment of this educational experiment integrates an interactive whiteboard, computer, projector, tablet PC, etc., where the tablet PC is one for each student (in line with the list of devices recommended by the North Teachers' University for the smart classroom environment). The interactive whiteboard is integrated with the interactive classroom platform, which contains a variety of common teaching toolkits, such as graphing tools (e.g., straightedge, semicircle, and coordinate system), brush tools, and baseboard tools (e.g., field grid, essay grid, and English alphabet grid), and has real-time synchronized screen dynamic operation, grouping, review, mutual evaluation, quizzing (all, group), randomly naming answers, voting, and student explanation. It also has functions such as sending and receiving assignments and recording screens. When the teacher initiates these learning activities, the students

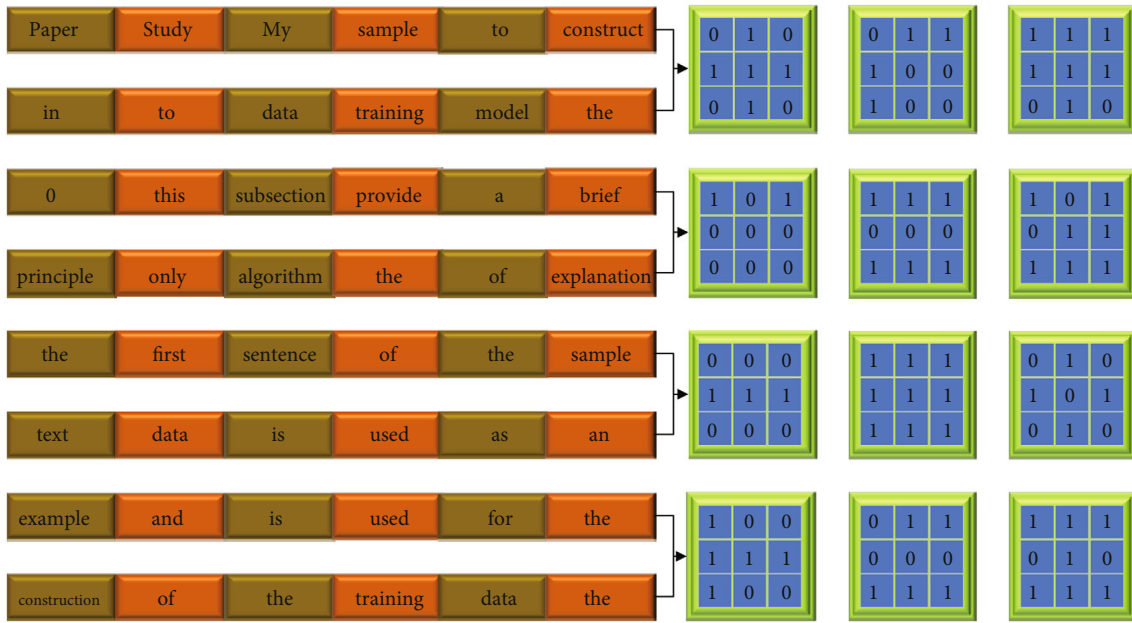


FIGURE 3: Construction process of training data.

respond on the tablet. These interactive features enable students to engage deeply in classroom learning, and in this experiment, they will also be used to support deep learning strategies such as lecture (both teacher-led and student-led), discussion, presentation, mutual evaluation, and critique [21]. Students can study at various kinds of free time. In addition, the interactive classroom platform can record students' answer processes to subjective questions, which is very helpful for in-depth analysis of students' cognitive and thinking processes. During the experiment, they will be used as materials for teachers or students to discuss, interassess, and critique.

In addition, students' learning time, learning progress, self-study and participation, microvideo viewing, cooperative and mutual learning questioning, assessment of learning, and reflection in this smart classroom can be recorded, analyzed, and fed back (mainly through the Creative Learning Platform). These data are presented to teachers and students in the form of visual reports, which provide them with the opportunity to grasp students' learning status and the situation in real time and provide data sources for teachers of this experiment to monitor students' learning status and achieve teacher data inspiration in decision flexibility. Compared to traditional machine learning algorithms, neural networks usually require more data, at least thousands of millions of labeled samples. In contrast, many machine learning problems can be solved well with less data if other algorithms are used. In addition, this smart classroom is equipped with six general whiteboards, two on either side of the interactive whiteboard (front side of the classroom) and the remaining four on both sides of the classroom, which provides conditions for this experiment to conduct multiple groups at the same time to carry out effective deep learning activities such as collaboration, inquiry, discussion, and presentation.

During class, teachers were asked to use an interactive whiteboard (integrated with the Interactive Classroom Plat-

form) for instruction, and students in both classes were asked to use personal tablets (integrated with the Creative Learning Platform) for learning. In both the experimental and control classes, the teacher facilitated student lectures (both teacher-led and student-led), discussions, presentations, mutual assessments, and reviews through the interactive classroom platform, and both classes provided approximately the same amount of time for students to conduct their investigations. The content is sent from the interactive classroom platform to the students' tablets [22]. During this time, students are supported by their tablets and the four whiteboards on either side of the classroom to engage in independent learning, group inquiry, collaborative work, and discussion and sharing. During this time, teachers act as supervisors, monitoring students' learning by walking around and viewing their learning activities and progress data in the Creativity Academy platform, determining whether individual students/groups need individual prompts or tutoring, moral support, and more difficult/easy learning tasks, and reviewing the appropriateness of the personalized learning tasks pushed by the platform. In addition, the teacher also acts as a guide and facilitator, providing tips, guidance, encouragement, or assistance to students with these needs, either through pushing them through the platform or by visiting them for consultation. This "interactive" flexible teaching model is shown in Figure 4.

According to the above teaching model, the teacher completes the deep learning teaching design in the Creative Learning Platform, in which the control class follows the original design and the experimental class follows the requirements of the deep learning sheet (see the subsection on intelligent classroom environment design) so that the concept of the deep learning sheet is integrated into the Creative Learning Platform of the experimental class, and it will play the role of a learning scaffold in the student learning process. This educational experiment is an exploration of

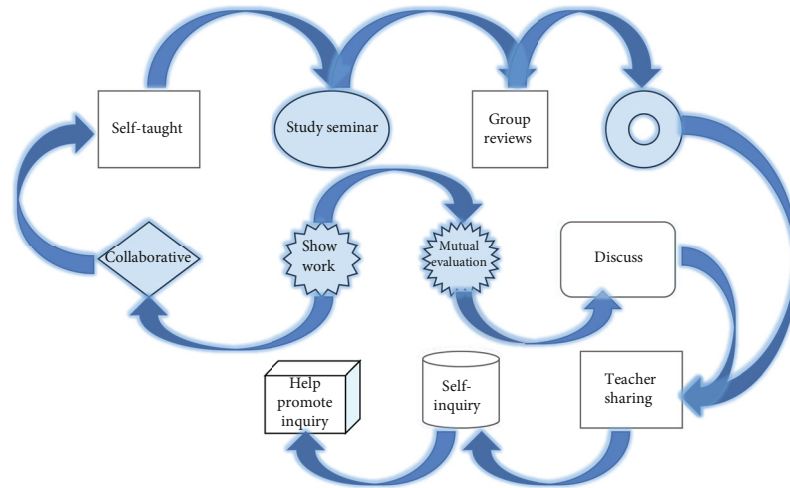


FIGURE 4: “Interactive” flexible teaching model.

the role of the scaffold, the learning sheet, in facilitating deep learning. Among other things, the SCS is an assessment of the quality of the smart classroom from the perspective of students’ perceptions. This is necessary because even if the hardware and software of the smart classroom environment are the same, students’ perceptions of it will be different, and such perceptions will influence the way students participate in the classroom and the learning strategies they adopt, thus directly or indirectly affecting their deep learning outcomes. This is because traditional text similarity algorithms, which focus on the similarity calculation of the text itself, therefore require a lot of normalization processes, such as normalization of numbers and Arabic numerals, and normalization of English units. Afterward, each student in both classes learned in the smart classroom environment of their respective classes. In both the experimental and control classes, the teacher presented the same learning objectives and topics and suggested reference materials in advance on the interactive whiteboard. The teacher briefed the students on these objectives, topics, and materials before the class began and followed the requirements in the Classroom Design above [23]. At the beginning of each independent inquiry, the teacher explicitly informed them that they could learn in any way they liked, but they could not leave the classroom, and students in the pilot class were additionally informed that they could complete the tasks on the in-depth learning sheet in the order they preferred (only those tasks required for this independent inquiry). Students were required to mark each task they completed in the learning process map (with frequent reminders from the teacher) to document their learning path.

The data collected by the SCS scale were analyzed by independent sample  $t$ -test and the results are shown in Figure 5. Overall, students in the experimental class perceived a higher level of quality in the smart classroom ( $M = 3.4884$ ,  $SD = 0.72121$ ) than those in the control class ( $M = 3.4703$ ,  $SD = 0.68382$ ). A closer look at the results of the ten-factor test shows (it is feasible to delve further into the dimensions of factors and domains<sup>2</sup>) that students in both classes perceived their smart classroom environment

to be not significantly different across the ten-factor dimensions (see Figure 5).

The key information extraction based on the text ranking algorithm achieves the purpose of reducing the document content and lays the foundation for the fuzzy identification of the content categories of academic documents in this subsection. The parsing of text data content can be generally divided into precise recognition and fuzzy recognition. Precise recognition is mainly applied to tasks with small text data size (generally not more than a few tens of characters) and fixed patterns of content to be recognized; fuzzy recognition is mainly applied to tasks with medium or large text data size and no fixed patterns of content to be recognized. For the research object and the research objective of this paper, the fuzzy recognition method based on edit distance will be used to intelligently identify the content categories of academic documents.

## 4. Analysis of Results

*4.1. Analysis of Identification Results.* Different document feature engineering models will be used to identify the filtered document dataset item by item according to the content categories of academic literature on cutting layout issues, and the text classification model evaluation system will be used to evaluate the test results and the model performance related to them. First, an ANN-based document feature model is used to identify the target type (O), and the results are shown in Figure 6 for journal 1 as an example (see Figure 6).

To evaluate the quality of the extracted relations, two sets of tests were conducted in this set of experiments: accuracy test and coverage test. To evaluate the accuracy of the extracted relations, we refer to the evaluation method of the YAGO system, and for each relation category, we randomly sample 200 relation tuples and check their accuracy manually. The coverage test is used to measure whether the extracted relations exist in the existing English knowledge graph. If not, it indicates that PNRE can extract new relations that cannot be covered by existing methods to complement the existing English knowledge graph. Within Broad point,

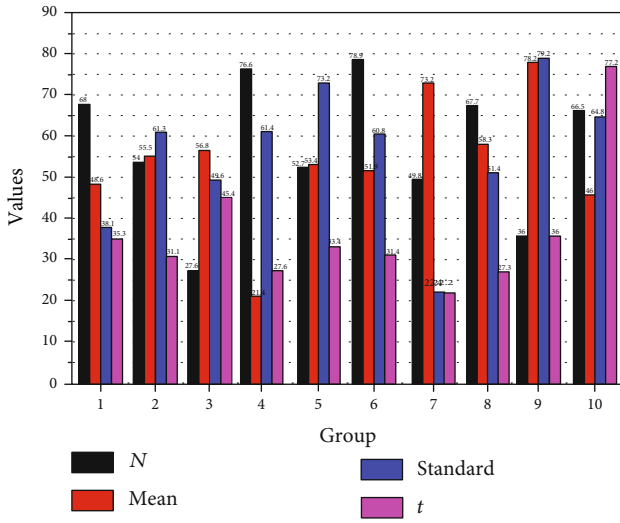


FIGURE 5: Descriptive statistics and independent sample  $t$ -test results.

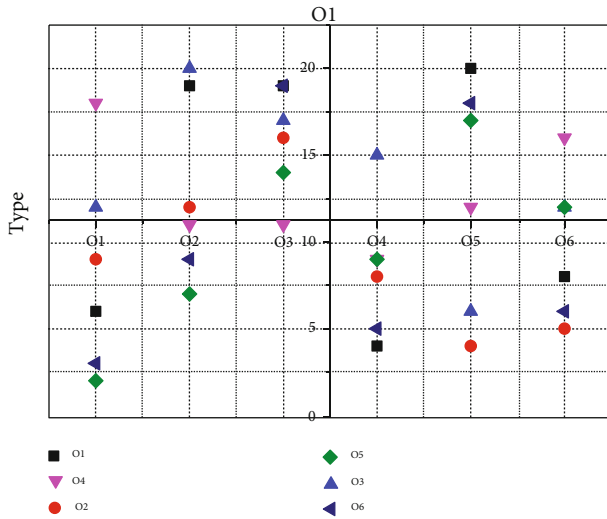


FIGURE 6: Confusion matrix of test results.

topic model has been successfully applied in many aspects, such as text classification features, relevance calculation, ctr prediction, precise ad targeting, and matrix decomposition. Specifically, based on the topic model, the topic distribution of texts and users can be calculated and used as features of pact and relevance, and it can also be used as a matrix decomposition method for dimensionality reduction, recommendation, etc.

When O1 and O2 are used as positive classes, all the performance indexes achieve good results, which indicates that the proposed method has a better recognition effect for the literature whose objective type is input minimization and output maximization; when O3 is used as a positive class, all the other three indexes achieve poor results except accuracy, which indicates that the proposed method has poor recognition effect for the literature whose objective type is multiobjective planning. To obtain the overall recognition performance of this paper's model for the target constraint,

a term to be recognized, the weighted average of each subclass is taken as the result, and the results show that this paper's method shows good overall performance for the recognition of the target constraint in journal 1. Following the same method, the target constraint recognition is tested for the literature in the remaining journals, and the results are shown in Figure 7.

The proposed method in this paper shows good performance in the accuracy of target constraint recognition of literature, and the overall recognition accuracy of each database and each journal is above 70%. When the journals are used as the test unit, the performance of the method in this paper is poor in recognizing the target constraints of the documents published in journals 3, 4, and 5. The worst results were found in journal 3, with the lowest values of 57.14%, 57.94%, and 55.37%, respectively. When the database is used as the test unit, the proposed method shows good recognition performance in the task of document target constraint recognition, and the results of each performance index of each database are above 75.00%, and most of the indexes are above 80.00%, which proves the effectiveness and feasibility of this method.

To prove that the research framework of this paper has the same effectiveness for the remaining document recognition tasks, this paper completed the testing work for all recognition tasks and produced the results. In this subsection, the recognition test results of the problem dimension and set features are shown, and the results are analyzed accordingly. From the above comparison results, due to the corresponding improvement of the text feature extraction model, the overall research scheme proposed in this paper achieves better results in each recognition task, and the weighted average value of the  $F1$  value of each database is above 70%, with the lowest being 70.17% and the highest being 86.78%, and all of them are significantly higher than the bag-of-words model and TF-IDF model, and the comprehensive acceptability. The overall acceptability was 81.19%, which reached the research goal of this paper.

**4.2. Analysis of Teaching Results.** According to the concept of the R-SPQ-2F scale, this effect of deep learning sheets on deep and superficial learning equations stems from both strategy means and motivation. As can be seen in Figure 8, the difference between the experimental and control classes in terms of deep strategy means (DS) was significant and had a moderate treatment effect level, and the difference in terms of deep motivation (DM) was also significant and close to a moderate effect level. Although the difference in the superficial learning strategy between these two classes was not significant, the differences between its two components were not insignificant; specifically, the difference between the experimental and control classes in the superficial strategy means (SS) was 0.892, which was not significant; however, the difference in superficial motivation (SM), which was significant ( $F(1, 84) = 4.470, p = 0.037 < 0.05$ ), had an actual effect value of 0.051, which was at the small effect level (see Figure 8).

The automatic construction of ontologies by libraries will save a lot of social costs and will be the focus of ontology

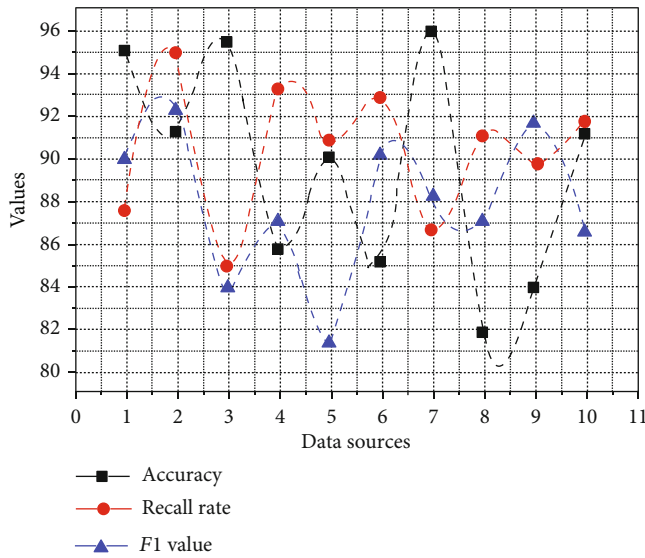


FIGURE 7: Target type recognition test results.

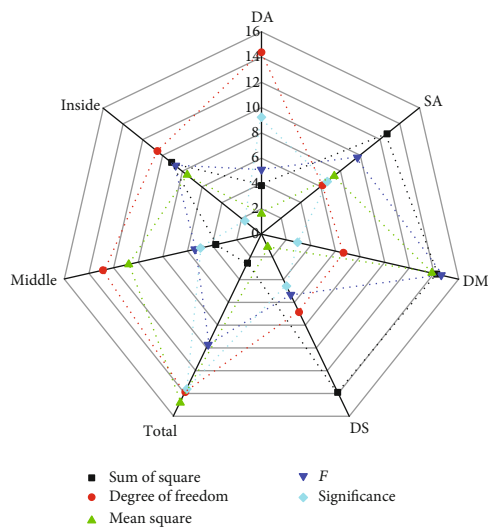


FIGURE 8: Results of one-way ANOVA for learning strategy.

construction now and in the future. In this paper, we summarize the mainstream ontology construction methods in the world today. In this paper, we summarize the mainstream automatic ontology construction methods in the world today and the main development directions of future ontology construction technologies are presented. To exclude the influence of prior creative experiences on students' creativity development, covariates were used as covariates in this substudy for the analysis of covariance. Because the K-DOCS scale, which is used to capture prior creative experiences, and the CDWR scale, which is used to assess creativity levels, use different scoring systems (i.e., they have different scales), they need to be standardized before the covariance analysis. Given the inability of linear transformations to eliminate the effect of the scale on covariance, this substudy used Z-scores to standardize the two types of data.

Due to the relatively small sample size of the experiment, the residuals of the previously created empirical data needed to be tested for chi-square and normal distribution after standardization to determine whether analysis of covariance could be performed. The Levene chi-square test was calculated to be  $F(1, 84) = 1.794, p = 0.184$ , and the Shapiro-Wilk test for normal distribution was  $p = 0.188$ , both of which were not significantly different ( $p > 0.05$ ), indicating that the experimental and control samples had homogeneous variances and were normally distributed, thus allowing for covariance between the previously created empirical data and the test data. In addition, due to the small sample size of this experiment, the actual effect sizes of the deep learning sheets need to be reported when reporting the results of the covariance analysis. With the support of the smart classroom environment, the deep learning sheets promote the development of deep learning skills (e.g., creativity) by increasing students' deep engagement in learning and guiding them to adopt deep learning strategies. Following this logic, this substudy constructs regression models with the smart classroom, classroom engagement, level of mindful flow, and deep learning strategies, which are the focus of the deep learning sheets, as independent variables, and the level of creativity as the dependent variable. Given the influence of prior creative experiences on creativity development, the present substudy also used them as independent variables in the multiple regression analysis.

### 5. Conclusion

Document key information extraction based on text ranking algorithm and text content fuzzy recognition based on edit distance is proposed. Due to a large amount of text data in academic literature, the adoption of the global matching recognition method will cause the problems of long solving time and low recognition efficiency, so this paper adopts a step-by-step recognition method: firstly, the key information is extracted from the original document, and secondly, the extracted information is matched for recognition of related items. Through the above method, the research objective of this paper is finally achieved. The overall research framework constructed in this paper was tested on the document dataset, the test results were analyzed in detail, and the analysis results verified the effectiveness of the method in this paper. Meanwhile, to make the research work of this paper have certain practicality, a prototype system of intelligent mining of academic documents is developed to realize the modularization, integration, and automation of the research methods in each stage.

### Data Availability

All information is within the paper.

### Conflicts of Interest


No competing interests exist concerning this study.

## References

- [1] S. Puri and S. P. Singh, "A fuzzy matching based image classification system for printed and handwritten text documents," *Journal of Information Technology Research*, vol. 13, no. 2, pp. 155–194, 2020.
- [2] F. Ren and Y. Bao, "A review on human-computer interaction and intelligent robots," *International Journal of Information Technology & Decision Making*, vol. 19, no. 1, pp. 5–47, 2020.
- [3] W. Liu, H. L. Ma, and A. Walsh, "Advance in photonic crystal solar cells," *Renewable and Sustainable Energy Reviews*, vol. 116, p. 109436, 2019.
- [4] X. Zhang, C. Zang, H. L. Ma, and Z. J. Wang, "Study on removing calcium carbonate plug from near wellbore by high-power ultrasonic treatment," *Ultrasonics Sonochemistry*, vol. 62, p. 104515, 2020.
- [5] H. L. Ma, X. Zhang, F. F. Ju, and S. B. Tsai, "A study on curing kinetics of nano-phase modified epoxy resin," *Scientific Reports*, vol. 8, no. 1, p. 3045, 2018.
- [6] K. I. Frank, M. Hibbard, M. Shucksmith et al., "Comparative rural planning cultures," *Planning Theory & Practice*, vol. 21, no. 5, pp. 769–795, 2020.
- [7] A. Charitopoulos, M. Rangoussi, and D. Koulouriotis, "On the use of soft computing methods in educational data mining and learning analytics research: a review of years 2010–2018," *International Journal of Artificial Intelligence in Education*, vol. 30, no. 3, pp. 371–430, 2020.
- [8] E. Raveh, Y. Ofek, R. Bekkerman, and H. Cohen, "Applying big data visualization to detect trends in 30 years of performance reports," *Evaluation*, vol. 26, no. 4, pp. 516–540, 2020.
- [9] M. Ling, M. J. Esfahani, H. Akbari, and A. Foroughi, "Effects of residence time and heating rate on gasification of petroleum residue," *Petroleum Science and Technology*, vol. 34, no. 22, pp. 1837–1840, 2016.
- [10] H. L. Ma and S. B. Tsai, "Design of research on performance of a new iridium coordination compound for the detection of  $Hg^{2+}$ ," *International Journal of Environmental Research and Public Health*, vol. 14, no. 10, p. 1232, 2017.
- [11] L. Y. Mo, W. H. Z. Sun, S. Jiang et al., "Removal of colloidal precipitation plugging with high-power ultrasound," *Ultrasonics Sonochemistry*, vol. 69, p. 105259, 2020.
- [12] D. Xu and H. Ma, "Degradation of rhodamine B in water by ultrasound-assisted  $TiO_2$  photocatalysis," *Journal of Cleaner Production*, vol. 313, article 127758, 2021.
- [13] M. Middell, "From universal history to transregional perspectives: the challenge of the cultural and spatial turn to world and global history in the 1970s and today," *Cultural History*, vol. 9, no. 2, pp. 241–264, 2020.
- [14] D. Levi and P. Tucker, "J after J. Ruskin: line in the art teaching of John Ruskin and Ebenezer Cooke," *Journal of Art History*, vol. 22, pp. 1–16, 2020.
- [15] D. Gao, Y. Liu, Z. Guo et al., "A study on optimization of CBM water drainage by well-test deconvolution in the early development stage," *Water*, vol. 10, no. 7, p. 929, 2018.
- [16] S. B. Tsai and H. Ma, "A research on preparation and application of the monolithic catalyst with interconnecting pore structure," *Scientific Reports*, vol. 8, no. 1, p. 16605, 2018.
- [17] J. Xie and H. Ma, "Application of improved APO algorithm in vulnerability assessment and reconstruction of microgrid," *IOP Conference Series: Earth and Environmental Science*, vol. 108, no. 5, article 052109, 2018.
- [18] K. Chakraborty, S. Bhattacharyya, and R. Bag, "A survey of sentiment analysis from social media data," *IEEE Transactions on Computational Social Systems*, vol. 7, no. 2, pp. 450–464, 2020.
- [19] J. Prüfer and P. Prüfer, "Data science for entrepreneurship research: studying demand dynamics for entrepreneurial skills in the Netherlands," *Small Business Economics*, vol. 55, no. 3, pp. 651–672, 2020.
- [20] T. Chaiechi, "Sustainable tropical cities: a scoping review of multidisciplinary methods for urban planning," *eTropic: Electronic Journal of Studies in the Tropics*, vol. 19, no. 2, pp. 25–51, 2020.
- [21] R. S. Liévanos, "Racialised uneven development and multiple exposure: sea-level rise and high-risk neighbourhoods in Stockton, CA," *Cambridge Journal of Regions, Economy and Society*, vol. 13, no. 2, pp. 381–404, 2020.
- [22] Z. K. Al, "Military and politics in Syria (1946-1963)," *Syria Studies*, vol. 12, no. 1, pp. 52–202, 2020.

## Research Article

# Online Data Migration Model and ID3 Algorithm in Sports Competition Action Data Mining Application

Li Ju <sup>1</sup>, Lei Huang,<sup>2</sup> and Sang-Bing Tsai <sup>3</sup>

<sup>1</sup>Department of Physical Education, Guangxi Medical University, Nanning, Guangxi, 530000, China

<sup>2</sup>Beijing Sport University, Beijing, 100000, China

<sup>3</sup>Regional Green Economy Development Research Center, School of Business, WUYI University, China

Correspondence should be addressed to Li Ju; juli12342010@sina.com and Sang-Bing Tsai; sangbing@hotmail.com

Received 2 June 2021; Revised 20 June 2021; Accepted 1 July 2021; Published 10 July 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Li Ju et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The ID3 algorithm is a key and important method in existing data mining, and its rules are simple and easy to understand and have high application value. If the decision tree algorithm is applied to the online data migration of sports competition actions, it can grasp the sports competition rules in the relationship between massive data to guide sports competition. This paper analyzes the application performance of the traditional ID3 algorithm in online data migration of sports competition actions; realizes the application steps and data processing process of the traditional ID3 algorithm, including original data collection, original data preprocessing, data preparation, constructing a decision tree, data mining, and making a comprehensive evaluation of the traditional ID3 algorithm; and clarifies the problems of the traditional ID3 algorithm. Mainly, the problems of missing attributes and overfitting are clarified, which provide directions for the subsequent algorithm optimization. Then, this paper proposes a  $k$ -nearest neighbor-based ID3 optimization algorithm, which selects values similar to  $k$ -nearest neighbors to fill in the missing values for the attribute missing problem of the traditional ID3 algorithm. Based on this, the improved algorithm is applied to the online data migration of sports competition actions, and the application effect is evaluated. The results show that the performance of the  $k$ -nearest neighbor-based ID3 optimization algorithm is significantly improved, and it can also solve the overfitting problem existing in the traditional ID3 algorithm. For the overall classification problem of six types of samples of travel patterns, the experimental data samples have the characteristics of high data quality, a considerable number of samples, and obvious sample differentiation. Therefore, this paper also uses the deep factorization machine algorithm based on deep learning to classify the six classes of travel patterns of sports competition action data using the previously extracted relevant features. The research in this paper provides a more accurate method and a higher-performance online data migration model for sports competition action data mining.

## 1. Introduction

Sports competition is a competitive recreational activity based on certain sports rules and is a very important form of sports activity. The viewing of sports competitions plays a unique role in meeting the exercise needs of people. Since their inception, sports competitions have attracted many participants and spectators with their unique charm [1]. However, when people study sports competitions, they tend to pay more attention to the technical factors and how to promote the level of competition, but few people study sports competitions from the aesthetic level, to explore the deep-

seated reasons why sports competitions are so attractive to people, and to develop a certain ideal space for sports competitions. In recent years, clustering technology has been developing, both in terms of clustering technology and clustering applications, and academics have made multifaceted research on clustering algorithms; recommendation algorithms have been applied to the Internet in recent years, and the increasingly diverse society has led to the rapid development of recommendation algorithms [2]. Applying data mining-related methods to sports competition action data mining can solve the situation in that the number of athletes is huge and the condition of the sport is complex and changeable,



which can effectively solve the problems faced by sports competition [3].

By comparing the previously learned knowledge with the new knowledge, we can find out whether there is a similarity between the two, and it is this similarity that forms the basis of transfer learning [4]. At present, it seems that such similarity exists in the vast majority of data, which allows migration learning to be widely applied to many areas of machine learning. Usually, the training of convolutional neural networks does not start from scratch, because as the size of the dataset increases, the time required to train the model is still long even with good hardware performance [5]. This avoids the problem of a long training time. The use of migration learning can mainly save the cost of model training, improve the efficiency of model training, and also optimize the final results. At the theoretical level, migration learning can be applied to any relevant domain with good results [6]. However, if the similarity between the models to be trained and the pretrained models is not good enough, the final results will not be good or even negative migration will occur, so the similarity between the models is the cornerstone for migration learning [7]. Clustering is a machine learning technique that involves the grouping of data points. Given a set of data points, we can use a clustering algorithm to divide each data point into a specific group. Theoretically, data points in the same group should have similar attributes and/or features, while data points in different groups should have highly different attributes and/or features. Clustering is an unsupervised learning method and is a common statistical data analysis technique used in many fields. In data science, we can use clustering analysis to gain some valuable insights from our data.

Currently, the amount of research on decision tree algorithms is increasing, and the focus of research is mostly on the improvement and optimization of decision tree algorithms to improve the classification accuracy of decision tree algorithms, the application effect of decision tree algorithms, and the improvement of the decision tree pruning process to improve the comprehensive application effect of decision tree algorithms. The machine learning method is a method in which a computer uses its data to derive a certain model and uses these models to predict the future [8]. This process is like the human learning process, except that machines can analyze large dimensional data and are tireless. The study is described as follows. Chapter one is an introduction, which first analyzes the background and significance of studying the application of decision tree techniques and transfer learning in sports competition action data mining, and provides an overview of the research content. Chapter two provides an in-depth analysis of the collected domestic and international research data and grasps the status of domestic and international research. Chapter three introduces the data processing and migration models, as well as the design study of data mining classifiers; evaluates the application effect of the optimized ID3 algorithm in sports competition action data mining; and proposes an application of sports competition action data mining based on the optimized decision tree algorithm according to the evaluation results. Chapter four analyzes the research results of this

paper and evaluates the overall effect of sports competition action data mining application. Chapter five is the conclusion and outlook, which summarizes the conclusions obtained by conducting this study and proposes further research directions and priorities for the shortcomings in the research conclusions.

## 2. Related Work

Clustering is an unsupervised machine learning approach that aggregates data items, observations, or feature vectors into groups. Currently, clustering techniques have been developed significantly, and as an important branch of data mining, clustering techniques have diverse applications in various fields [9]. Karmani et al. proposed a maximum edge clustering (MMC) method, which is based on the support vector machine model. The results of the study proved that the clustering results are better than the  $K$ -Means algorithm [10]. In web data mining, structured text processing with structural and semantic coincidences is a challenge in the field of data mining, and processing XML document data is a challenge. Ma and Tsai used the tree meta-ancestor approach to identify XML document data and use clustering methods to process these data and make experiments accordingly [11]. When the number of athletes increases, it is not possible to effectively personalize the decision based on individual circumstances, so the efficiency is low and the response strategy and results are not satisfactory. In terms of data mining, we can automate the modeling process and use a series of techniques such as clustering and recommendation algorithms to help experts make recommendations and give coping strategies for different aspects of athletes, which is rarely done in sports competitions [12].

With the in-depth study of sports competition, the study of the ideas and theories of sports competition has penetrated all fields of sports. Gao et al. analyzed basketball comprehensively from two perspectives, the core level and the auxiliary level, which involved many aspects of basketball such as the body, technique, tactics, stars, confrontation, spirit, style, and the beauty of form, costume, and field equipment, which have positive significance for the study of the characteristics of basketball sports competition [13]. When the number of layers of neural networks increases, the learning ability of neural networks does not improve but, on the contrary, may become worse, which was later proven to be due to the disappearance of gradients [14]. Xie and Ma proposed the SVM algorithm, which became the mainstream of machine learning algorithms at that time because of its excellent migration performance that demonstrated greater advantages over neural networks in many problems [15]. Fujin et al. proposed the convolutional neural network (CNN), which is the first deep learning algorithm with the practical significance of a multilayer network structure [16]. It uses the spatial relativity of data and a human visual neural structure to reduce the number of network parameters, thus effectively improving the performance of the model [17]. A good data preprocessing method not only eliminates structural defects in the existing data but also prepares the data for data mining. The period of archery movements among individuals is not equal, and there are differences in the

timing of applying movement techniques at each stage. Equal-width discretization is suitable for handling complex data with chaotic data structure and strong data continuity [18].

Data mining algorithms such as association rules, clustering, and Markov-based data mining algorithms have been intensively studied and applied in sports competition action data mining to achieve the set research objectives. Several specialized data mining tools have been developed for different domains. The diversity of data mining tasks determines that data mining faces many challenging topics [19]. For data mining researchers, designing data mining languages, adopting efficient and useful data mining methods, and developing systems to build interactive and integrated data-mining environments are the main issues. Processing the collected kinematic data of sports competitions into a standard data structure suitable for use in data mining techniques is a major challenge. Since the speed of different athletes' sports movements varies, the preprocessing of the data is necessary to ensure the integrity of the data and the structure of the data. How to adapt kinematic data to data mining techniques requires a lot of experimentation and detailed analysis. We analyze the strengths and weaknesses of the models studied in the relevant references and use them to determine our own research model.

### 3. Online Data Migration Study of Sports Competition Action Based on ID3 Algorithm

**3.1. Data Processing.** A suitable data model is built based on the dataset and the meaning it is intended to convey. The data being analyzed is the training set, and each attribute corresponds to a class label. The samples are randomly selected from the training subset samples divided by the training dataset, and the data are classified under the guidance of the class label number, which is often referred to as guided learning. The concept of guided learning is that this data analysis knows exactly which class is guiding the training samples, as opposed to not knowing which class is guiding the training samples, unguided learning, or what is called clustering. Data analysis is performed on the decision tree that has been constructed. There are many ways to perform data analysis on decision trees; generally, the accuracy of the decision tree is analyzed first, because if a decision tree cannot guarantee its accuracy, the other data analysis loses its meaning. Therefore, the accuracy rate indicates the proportion of data that can be correctly classified by the decision tree classification algorithm and is the most important measure. We generally use the known information rules as the standard and compare the results inferred from the model constructed with the decision tree with the standard [20]. If the comparison results are very different, the accuracy rate is lower, and if the difference is smaller, the accuracy rate is higher.

Data mining, also known as database file expertise discovery (KDD), is the noncommon process of obtaining reasonable, novel, potentially effective, and finally understandable ways from a lot of data. Preprocessing is an important component of the association analysis of datasets. Noise elimination,

data synthesis, and data standard unification are performed on the dataset using preprocessing to facilitate the subsequent data management analysis. In general, the specific requirements of preprocessing are the following: first, the smoothing of the processed dataset, mainly to eliminate the noise in the data, using box division, clustering, and query techniques; second, the synthesis of the dataset, using a concept hierarchy, replacing the low-level "raw" data with high-level concepts; and third, ensuring the standardization of the data—scaling attribute data into a small specific interval [21].

Let the set of datasets be  $D$  and the data samples in the dataset  $D$  be  $d$ . Assume that the dataset  $D$  has  $m$  different class attributes with different values, which are labeled as  $B_i$  ( $i = 1, 2 \dots n$ ). Therefore, the amount of classification information can be expressed by equation (1), where the weight of class attribute  $B_i$  is expressed by  $b_i$ , which can be calculated by  $b_i/d$ . The logarithmic function with a base of 10 is used here because the information is encoded in binary.

$$f(D) = \sum_{i=0}^n b_i \ln b_i. \quad (1)$$

Suppose that one of the attributes, here denoted as  $A$ , has  $v$  different values, which can be denoted as  $\{b_1, b_2 \dots b_m\}$ . Thus, the dataset  $D$  can be divided into  $v$  different subsets by the attribute  $A$ , here denoted as  $\{D_1, D_2 \dots D_m\}$ , where  $D_j$  ( $j = 1, 2 \dots m$ ) is denoted as the set of samples with the same value  $b_j$  ( $j = 1, 2 \dots n$ ) on the attribute  $A$  set. Suppose  $b_{ij}$  denotes the total number of samples belonging to category  $B_i$  in subset  $D_j$ . The formula for calculating the information entropy of attribute  $A$  is shown in (2), and  $f(x)$  denotes the proportion of the number of samples with the value  $b_j$  to the total number of samples for attribute  $A$ .

$$\begin{cases} f(x) = \frac{\sum_i^n b_{ij}}{d}, \\ G(D) = \sum_{j=1}^m f(x)_j f\left(\sum_{i=1}^n b_{ij}\right). \end{cases} \quad (2)$$

The information gain of attribute  $A$  can be calculated from

$$H(A) = G(D) - G(D_A). \quad (3)$$

Deep learning is proposed in response to the shallow learning phenomenon of mechanical learning, rote learning, and knowing what is right but not knowing what is wrong in practice. The "depth" here refers to the depth of learning of students. We do not require teachers to adopt a fixed model or method but rather emphasize that teachers should use appropriate methods to trigger, promote, and enhance students' deep learning. In this sense, deep learning is the opposite of shallow learning, and it is a criticism of the times.

The strategy of the decision tree classification algorithm is to calculate the information gain rate of all the test attributes in the alternative dataset and to use the attribute with

the maximum information gain rate as the current division attribute and finally to complete the construction of the decision tree by iterating the above process. The data were analyzed using divisional clustering, but the results were found to be unsatisfactory and not well differentiated [22]. Therefore, when conducting the reason analysis, the small amount of data and many data dimensions may have an impact on the clustering results, and the improved algorithm was considered according to the characteristics of sports competition action data. The improved  $K$ -Means algorithm is applied to the sports competition action data to mine the clustering results. The normalization formula is shown in equation (4), where  $\max$  and  $\min$  sub another and  $J$  is the maximum and minimum values of the sample data.

$$F(x) = \frac{x - \min(G(D))}{\max(G(D)) - \min(G(D))}. \quad (4)$$

Clustering all attributes of the data makes the clustering results reflect the relationship between metadata more fully; on the other hand, in cluster analysis, the attributes of the tuple inevitably form high attribute clusters with similarity. Due to the attributes of all data, clusters have the characteristics of clustering, and each group of data within its data cluster contains attributes; the formation of attribute clusters on each attribute will inevitably occur simultaneously. In other words, if the number of tuples in the dataset presatisfies the minimum support, then the cluster formed by all attribute clusters becomes the maximum item frequency set of the data cluster [23].

**3.2. Online Data Migration Model Construction.** To address the problem of changing the cost of online data, this paper proposes an adaptive cost-sensitive online migration learning method. First, we introduce the marker distribution into the traditional hinge loss function to calculate the classification cost adaptively; second, we combine the source and target domains using the combination parameters to realize the online migration from the source to the target domain; finally, we migrate to learn the classification model based on both cost and accuracy [24]. For each sample, we automatically calculate the adaptive cost of that sample based on the ratio of the current sample polarity to the polarity of all samples and then add the adaptive cost to the fusion to the loss function and use it to iteratively update the classification model to obtain the current latest classification model. The online data migration model in this paper is shown in Figure 1.

When the sample  $m_x$  is read, the positive and negative ratios in the sample set at the previous  $m$  moments will change, which will inevitably cause the final classification model to have a learning bias for some samples. To reduce the impact of this change, the current ratio of positive and negative samples and the ratio of the loss of the current sample to the total loss are multiplied together and denoted as  $D_m$ , as in equation (5). Where  $H_m^+$  and  $H_m^-$  represent the number of positive and negative samples, respectively,  $g(b_m, y_m, f(x_m))$  is the cost hinge loss function with introduced cost and  $g(D_m, b_m, y_m, f(x_m))$  is the loss function with introduced marker distribution parameters.

$$Y(m) = \frac{H_m^+}{H_m^-} * \frac{g(b_m, y_m, f(x_m))}{\sum_{i=0}^m g(D_m, b_m, y_m, f(x_m))} * 100\%. \quad (5)$$

Combining multiple weak classifiers using weight parameters to obtain a strong classifier is the  $\beta$  common approach. The adaptive cost-based online migration algorithm similarly combines the initial classifier  $f(0)$  and the online adaptation function  $w(m_x)$  by combining the parameters  $\beta$  to obtain a combined classifier.

$$\text{Classifier} = \beta f(0) + (1 - \beta)w(m_x). \quad (6)$$

In the online migration learning process, we aim to update  $m_{x-1}$  to  $m_x$  using a suitable algorithm. The classification model is obtained based on both smoothing and minimizing the cost sum. The distance between  $m_{x-1}$  and  $m_x$  is smoothed using Euclidean distance as in

$$M(x) = \min \left( \frac{(m_x - m_{x-1})^2}{2} \right), \quad x \subseteq [0, m]. \quad (7)$$

The cost changes as the online data changes, so it is necessary to find a way to make the cost change adaptively. The misclassification cost is updated adaptively using the ratio of positive and negative samples, and the updated cost can be used to dynamically adjust the learning of the classifier for different samples. Since the cost of rare samples is usually high, a classifier based on adaptive cost tends to improve the classification accuracy of rare samples, which in turn improves the performance of classification.

The availability and prevalence of large amounts of data and the use of tools to properly extract knowledge information have become very frequent. This fact has changed the traditional data analysis by orienting the data to certain specialized techniques under data science. In short, data science can be considered as a discipline that discovers new and important relationships, patterns, and trends after examining large amounts of data. Thus, data science techniques pursue the automatic discovery of knowledge contained in information stored in large databases.

The structural similarity matrix is used to describe the distribution between a descriptive attribute and a categorical attribute; the number of occurrences of all attribute values of each descriptive attribute under all attribute values of the categorical attribute and the attribute values of the categorical attribute are used to obtain the values with the highest number of overlaps to participate in the next step. The structural similarity matrix model will create a matrix for each described attribute to provide the corresponding parameters for calculating the sample structural similarity.

Since the sample structural similarity of the attributes is calculated based on the structural similarity matrix, we can assume that if the structural similarity matrix is verified to be free of multivalued bias, we can conclude that there is no multivalued bias through the sample structural similarity of the attributes [24]. And if it can be verified that there is no definite size relationship between the structural similarity matrix and the increase of attribute values and since the

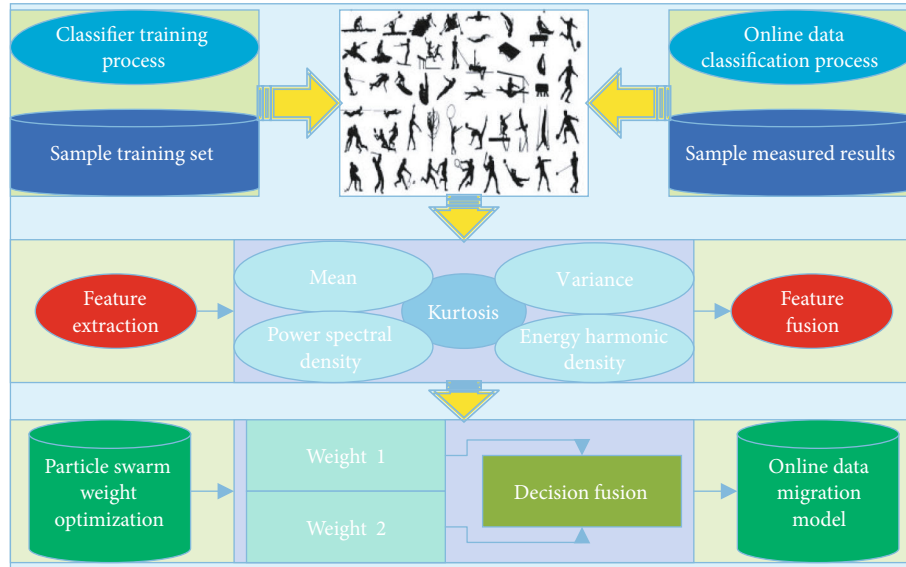


FIGURE 1: Online data migration model.

sample structural similarity generated by the structural similarity matrix is more biased towards the descriptive attributes that are structurally more similar to the categorical attributes, then the sample structural similarity as weights will play a positive role in the correction of information gain, thus reducing the interference of the multivalued bias problem on the selection of the joints.

**3.3. Data Mining Classifier Design.** In this paper, we propose an automatic extraction strategy for sports competition action data based on image overlap region feature migration machine learning, as shown in Figure 2. The strategy can be divided into three steps: (1) select the classifier with the highest generalization ability among ID3 algorithm classifiers to supervise the data classification and set it as the source classifier model; (2) propagate the source marker to the adjacent target image overlap region, randomly select a certain proportion of its marker samples to mix with the source training samples to form the pseudosample, and use the migration learning model to balance the difference of its temporal distribution; and (3) obtain the final classification model through the source classifier. The final classification model is obtained after continuous iterations, and the distribution difference between adjacent target image data is continuously reduced due to the increasing information of balanced temporal samples, to achieve the purpose of accurate extraction of sports competition action data (see Figure 2).

Using the automatic sports action data extraction strategy, only the source training samples of the source images and the classified images with overlapping areas with the images to be classified are involved in the classification, and then, the temporal difference balance based on the migration learning model and the source classifier model can complete the classification, and the information extraction of each image is completed one by one through the continuous iterative update to realize the large area sports competition data extraction. The action data extraction work is completed

one by one through continuous iterative updates. In this paper, the classifier algorithm with high adaptability and generalization ability is used as the classification model, which can effectively improve the accuracy of image classification results by learning more potential features. The migration learning model, as a tool for adaptive balancing between source and target images, can reduce the differences in data distribution between them due to temporal and spatial differences and has a weakening effect on improving the problem of obvious color differences in the extraction of thematic information covering large regions of sports competition action data. Selecting certain marker samples in the overlapping area together with the source training samples constitutes a pseudosample, which makes it possible to consider the spectral feature distributions of both the source and target images when performing temporal phase difference balancing and avoid the overbalancing situation. The close cooperation of the above methods to complete the large area grass information extraction will help to improve the extraction accuracy.

## 4. Analysis of Results

**4.1. ID3 Algorithm Performance Analysis.** Comparing the running time of the traditional ID3 algorithm and the improved ID3 algorithm, the execution time of the improved ID3 algorithm is shorter and the running time is reduced by the improvement, and the comparison results are shown in Figure 3(a). By comparing the classification error rate of the traditional ID3 algorithm and the improved low algorithm, the error rate of the improved ID3 algorithm is lower, and the comparison results are shown in Figure 3(b). By comparing the experiments of both the traditional ID3 algorithm and the improved ID3 algorithm in terms of algorithm execution time and classification error rate, it can be found that the dataset has reduced running time and lowered the

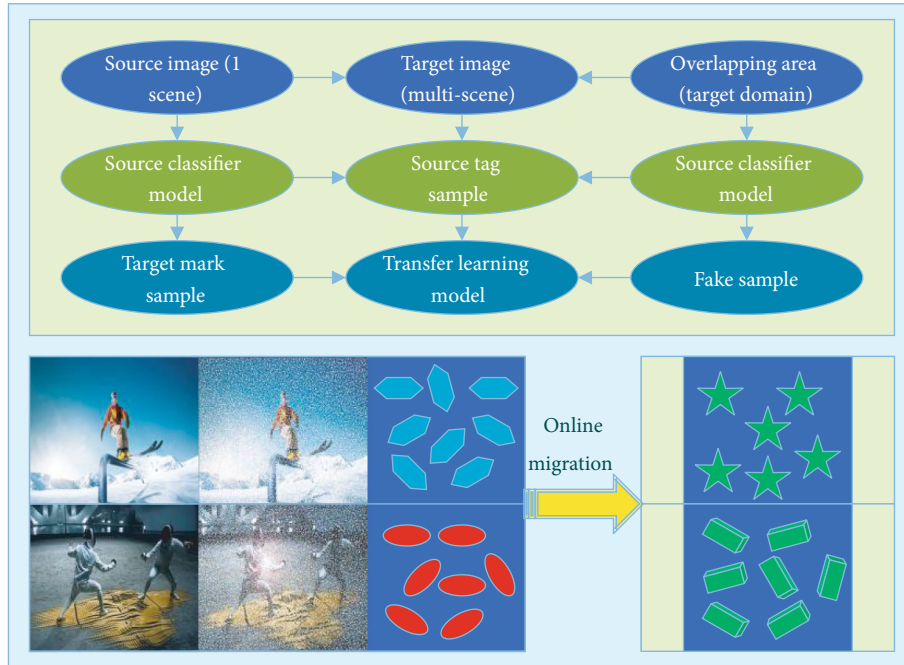
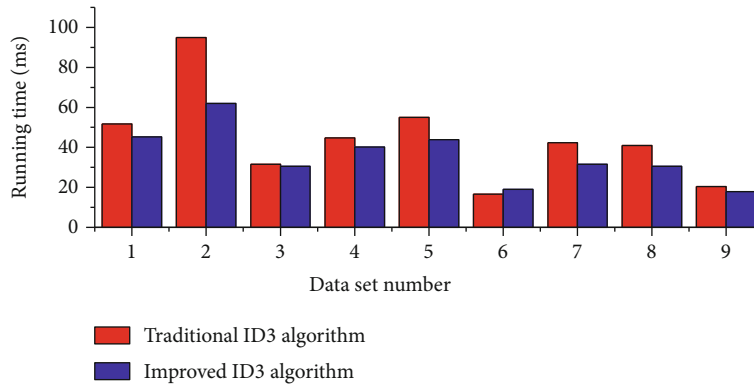
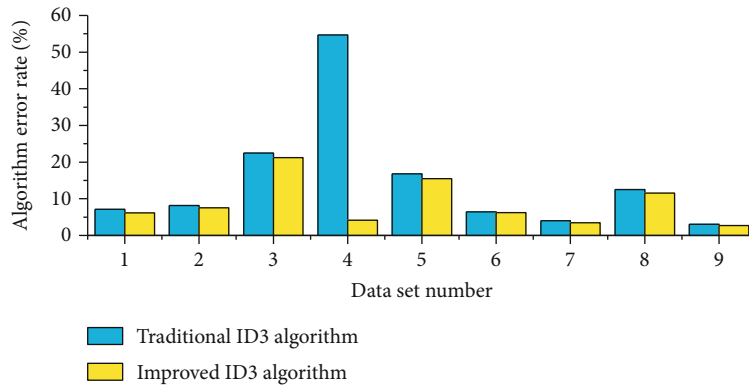


FIGURE 2: Flowchart of automatic extraction strategy for sports competition action data.



(a)



(b)

FIGURE 3: Comparison of the running time and error rate of the algorithm.

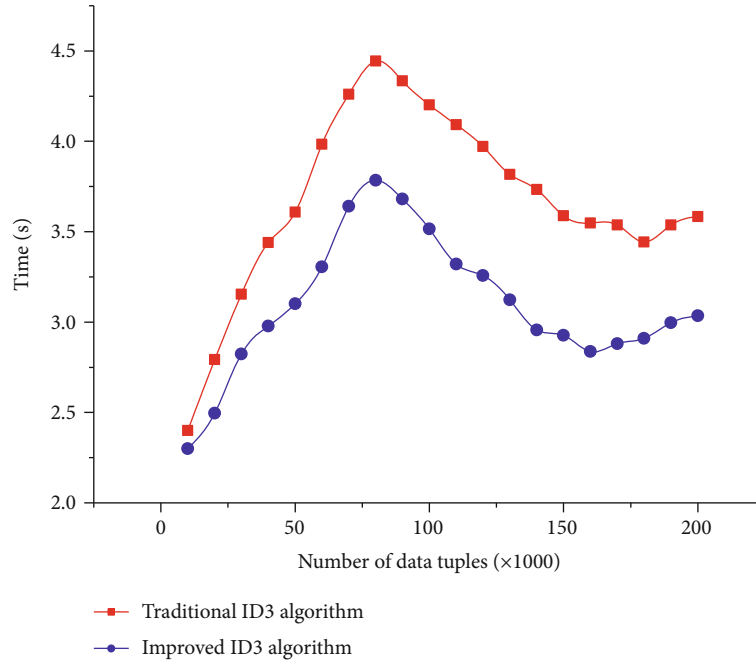


FIGURE 4: Modeling time comparison results.

average classification error rate after using the improved ID3 algorithm (see Figure 3).

Although the improved ID3 algorithm is proportional in time cost to the number of data tuples in the sample dataset, as the number of data tuples increases, this ratio is converging to the time to quantity ratio of the original algorithm. This means that the larger the sample dataset, the closer the decision tree algorithm is to the traditional ID3 algorithm in terms of practical performance. Then, we continue to process the time information to get the results of modeling time comparison between the traditional ID3 algorithm and the improved ID3 algorithm (Figure 4). Figure 4 shows that the time ratio between the improved ID3 algorithm and the traditional ID3 algorithm increases when the number of data tuples in the sample dataset grows from 10,000 to 20,000, while the time ratio between the improved ID3 algorithm and the traditional ID3 algorithm continues to decrease when the number of data tuples grows from 20,000 to 80,000, but at the value of 80,000, the decreasing trend of this ratio slows down at 160,000 items. This trend indicates that the time efficiency of the improved ID3 algorithm decreases when the number of data tuples is small, mainly due to the more complex algorithm, but the proportion of time spent by the algorithm decreases when the number of entries increases steadily. When the number of data tuples in the sample dataset is very large, the main time overhead has shifted to data I/O and processing and is no longer concentrated on the algorithm (see Figure 4).

The sample structural similarity model used by the improved ID3 algorithm does not have a large impact on the classification performance for datasets with an overall low structural similarity. When encountering datasets where the number of attribute values does not differ much and the structural similarity of the descriptive attributes is not more

obviously related to the structural similarity of the classified attributes in terms of structure, the structural similarity has less impact on the generation of the final decision tree structure, thus making the prediction rate of the improved ID3 algorithm more approximate to that of the traditional ID3 algorithm and making the classification accuracy between the two closer.

*4.2. Online Data Migration Model Analysis.* Figure 5 shows the experimental results of the VLSC dataset, and the following conclusions are obtained from the figure: (1) All migration learning methods perform better than online methods, the results indicate that the effect of training using only the target sample is limited, and using source samples with the same source or the same structure as the target sample is beneficial to the learning of the target sample. (2) The classification effect of kernel-based migration methods is significantly higher than that of online algorithms. Relative to the classical online algorithm, the classification accuracy of the kernel-based migration method is about 10% higher in the six tasks generated by the sets Caltech101 and SUN09 and exceeds 5% in the tasks of the datasets Label ME and VOC2009, which demonstrates the applicability of the kernel-based migration learning method. (3) Compared with the basic kernel-free online migration method, the overall accuracy of KOTL experiments is improved by about 2%, which indicates that the kernel-free online algorithm is suitable for the case of linearly separable source and target domains, while the kernel-based online migration method solves the problem and is also suitable for the case of linearly separable domains. (4) For the offline kernel migration method ARRLS, KOTL is also comparable to it and applicable to the online case. The offline kernel migration method requires a large storage space for caching all samples in both source and target

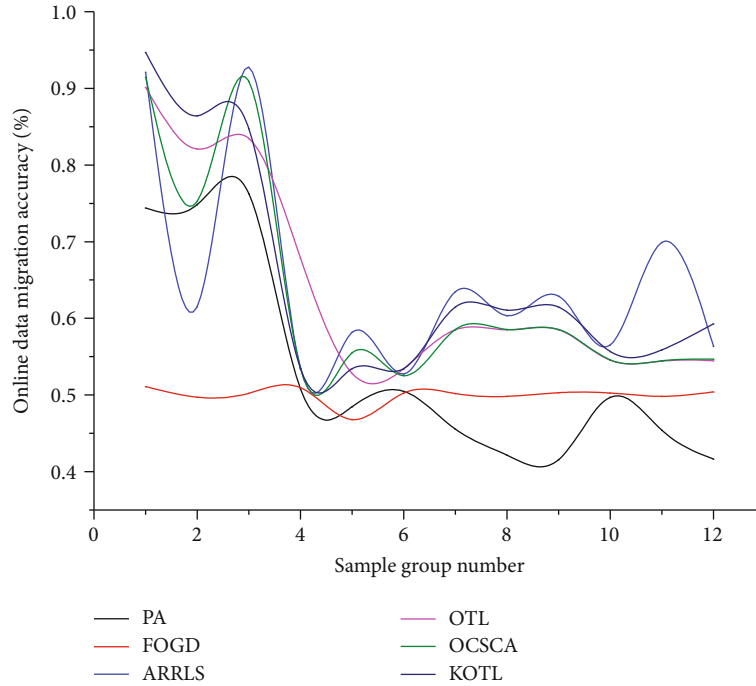


FIGURE 5: Online data migration accuracy comparison results.

domains, and the computational effort is also related to the number of samples (see Figure 5).

To verify the migration effect of the online data migration model proposed in this paper, as well as the enhancement effect on the target domain data under different sparsity, maintain the consistent time interval between samples, and maintain the temporal order of the data, this experiment selects the first 10%~90% samples from SCADA data as the target domain data and the remaining data as the test set for power prediction, using a long short-term memory network (LSTM) as the regression model for wind power prediction, and root mean square error (RMSE) is used as the measure. The root mean square error (RMSE) is the square root of the ratio of the square of the deviation of the predicted value from the true value to the number of observations  $n$ . In practical measurements, the number of observations  $n$  is always limited and the true value can only be replaced by the most trustworthy (best) value. The RMSE results of the prediction results on the test set are shown in Figure 6 for the comparison between the online data migration model and the PCTR model and the case where no source domain data are introduced. As the proportion of target domain data decreases and the data sparsity of the source domain training data increases, the gap between the prediction accuracy of the nonmigrated model and the migrated model becomes more obvious. The experimental results show that the online data migration model can effectively solve the problem of sparse target domain training samples and improve the accuracy of the prediction task by introducing source domain data and migrating the source domain knowledge (see Figure 6).

4.3. Analysis of the Effect of Sports Competition Action Data Mining. We have added the validation set test set. For each

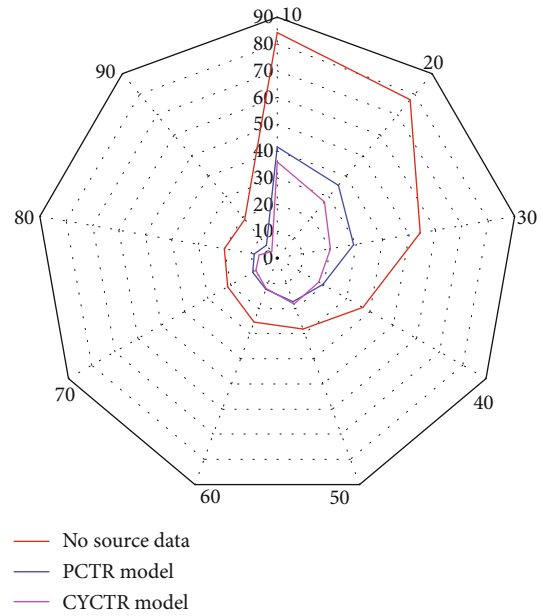


FIGURE 6: RMSE values of the model on the test set.

competitive sports development situation, input indicators, output indicators, and development indicators are used as indicator selection basis. Among them, the expenditure on competitive sports (including sports competition fees, sports training fees, and sports stadium construction fees) is taken as the input indicator, the points of the sports competition are taken as the output indicator, and the number of outstanding athletes is taken as the development indicator. Figure 7 shows the descriptive statistics about the three initial variables, including the mean, standard deviation, and the

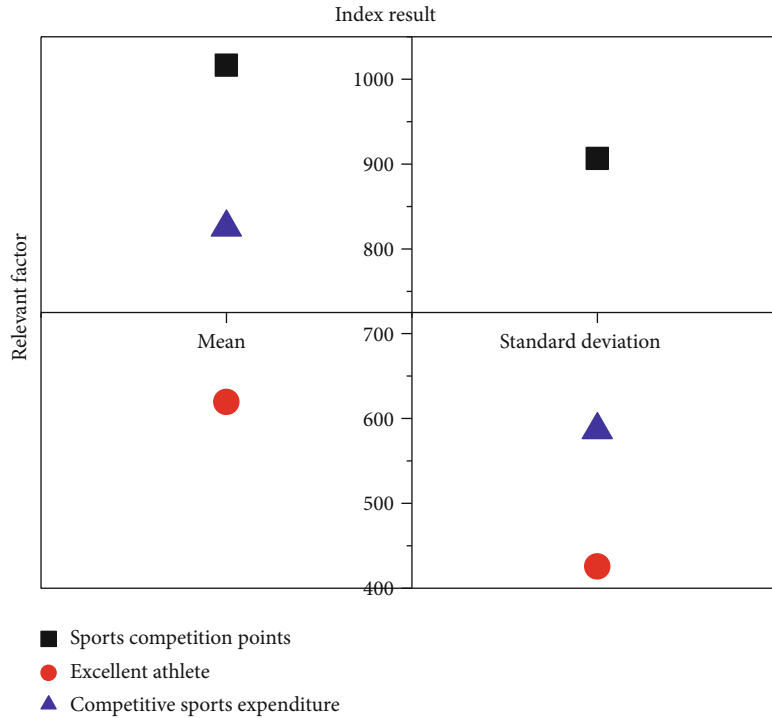


FIGURE 7: Descriptive statistics of the initial variables.

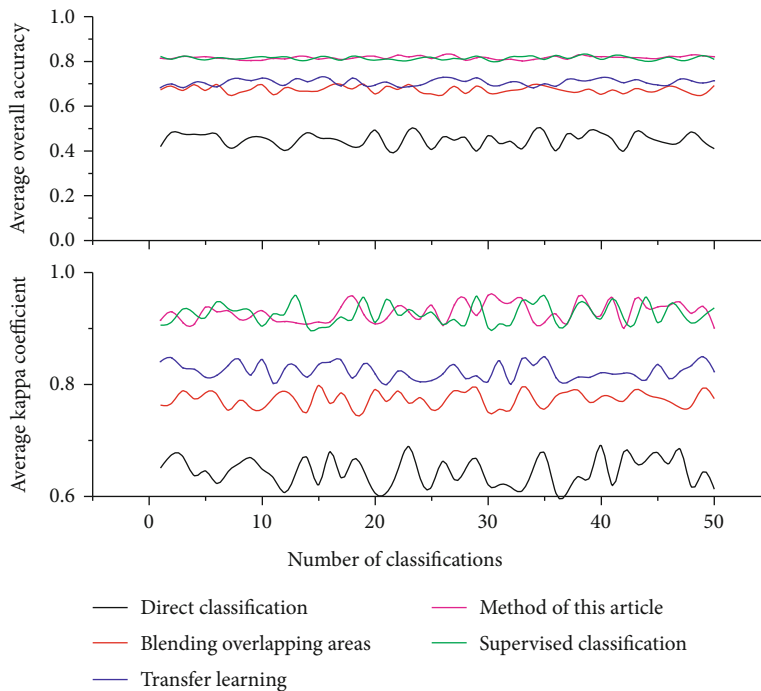


FIGURE 8: Comparison of the overall classification accuracy of the five methods.

number of values used in the analysis  $N$ . Among them, the standard deviations of the points of sports competitions, excellent athletes, and competitive sports expenditures are 907.56, 426.47, and 586.61, respectively (see Figure 7).

Figure 8 shows the comparison of the average overall accuracy and average kappa coefficient of the target

images under the five classification strategies. With the overall accuracy as the evaluation index, the relationship of  $(a) < (c) < (b) < (d) \approx (e)$ , and the kappa coefficient as the evaluation index, the relationship of  $(a) < (b) < (c) < (e) < (d)$  is satisfied. The overall classification situation still confirms the above conclusion that both fusion of overlapping



region labeled sample information and migration learning can improve the image classification accuracy, while the fusion of overlapping region labeled sample information and migration learning can add up the ability of both to improve the classification accuracy. The average overall accuracy obtained by this method is slightly equal to that of the supervised classification, the average kappa coefficient is greater than that of the supervised classification, and the two classification accuracy evaluation metrics are given the same weight, which indicates that this method is slightly better than the supervised classification method (see Figure 8).

When performing data mining algorithms, the sports competition dataset is used as the research object for the online migration success rate of sports competition action data, so the decision tree algorithm is used to analyze the main attributes in the sports competition data source that may affect the check-in rate and to identify the factors that are most likely to affect the sports competition movement so that these potentially relevant influencing factors can be scientifically applied to provide key future sports competition decision basis. The advantages of the decision tree algorithm are simple and easy to understand analysis, high classification accuracy, and high execution efficiency, so the decision tree algorithm is suitable to be applied to massive data mining. The ID3 algorithm selects the classification criterion of information gain, it selects the attribute with more attribute values as the split attribute, and the ID3 algorithm can only mine the data with nonlinearity.

Data classification is an important part of data mining, and the decision tree algorithm is one of them. It is an important method in data classification methods and has a mature theoretical foundation and a good development platform. The decision tree algorithm is relatively easy to understand, and the construction process is fast. What is more, decision trees can be easily converted into SQL statements for efficient access to the large number of various databases existing in financial systems. Data mining technology is an important technical tool to transform data into knowledge and value, but traditional data mining technology faces many challenges to extract the rich knowledge and value implied from big data. An important way to solve the problem of big data mining is to research and develop more efficient data mining algorithms based on the essential characteristics of big data.

## 5. Conclusion

In this paper, the clustering algorithm in data mining is applied to the sports competition action data migration model; using the theory related to the ID3 clustering algorithm and combining the characteristics of sports competition action data, the ID3 algorithm is combined and the ID3 algorithm is slightly improved to make it more suitable for sports competition action number nuggets. Then, the data of each dimension are clustered to derive the corresponding results, and the results are synthesized and analyzed to facilitate the subsequent recommendation algorithm for online data migration modeling. This paper presents a detailed analysis of sports competition action data and related theories and successfully applies the theoretical

knowledge of computer aspects to sports competition action data mining, paving the way for future cross-application of sports competition action analysis and computer disciplines. In this paper, only the ID3 algorithm in the data mining clustering algorithm has been fully studied and other clustering algorithms are not involved, and the algorithm based on division requires more data types and shapes of datasets and is also easy to fall into the local optimum; whether the sports competition pressure data in this paper is coincidentally applicable to other algorithms is not studied in-depth, and other types of clustering algorithms can be used for the data in this paper subsequently. It is possible to reach better conclusions. The clustering results have some errors and are only provided as a reference to the also-ran scholars for reference and cannot replace the actual also-ran judgments. Similarly, the recommendation results only assist the scholars to make suggestions and cannot be provided to the athletes as actual suggestions, but this, as a practical nature of the exploration, still needs continued in-depth research in the follow-up work. We provide an in-depth analysis of the online data migration model and the ID3 algorithm in sports game action data mining, and the model under study has a strong practicality.

## Data Availability

All information is within the paper.

## Conflicts of Interest

No competing interests exist concerning this study.

## References

- [1] S. Chen, "An effective going concern prediction model for the sustainability of enterprises and capital market development," *Applied Economics*, vol. 51, no. 31, pp. 3376–3388, 2019.
- [2] B. V. Chowdary and Y. Radhika, "A survey on applications of data mining techniques," *International Journal of Applied Engineering Research*, vol. 13, no. 7, pp. 5384–5392, 2018.
- [3] J. M. Torres, C. I. Comesaña, and P. J. Garcia-Nieto, "Review: machine learning techniques applied to cybersecurity," *International Journal of Machine Learning and Cybernetics*, vol. 10, no. 10, pp. 2823–2836, 2019.
- [4] S. Shafqat, S. Kishwer, R. U. Rasool, J. Qadir, T. Amjad, and H. F. Ahmad, "Big data analytics enhanced healthcare systems: a review," *The Journal of Supercomputing*, vol. 76, no. 3, pp. 1754–1799, 2020.
- [5] D. Draskovic, M. Cvetanovic, and B. Nikolic, "SAIL—software system for learning AI algorithms," *Computer Applications in Engineering Education*, vol. 26, no. 5, pp. 1195–1216, 2018.
- [6] C. Schmidt and W. N. Sun, "Synthesizing agile and knowledge discovery: case study results," *Journal of Computer Information Systems*, vol. 58, no. 2, pp. 142–150, 2018.
- [7] W. Liu, H. L. Ma, and A. Walsh, "Advance in photonic crystal solar cells," *Renewable and Sustainable Energy Reviews*, vol. 116, article 109436, 2019.
- [8] X. Zhang, C. Zang, H. L. Ma, and Z. J. Wang, "Study on removing calcium carbonate plug from near wellbore by high-power

- ultrasonic treatment,” *Ultrasonics Sonochemistry*, vol. 62, article 104515, 2020.
- [9] H. L. Ma, X. Zhang, F. F. Ju, and S. B. Tsai, “A study on curing kinetics of nano-phase modified epoxy resin,” *Scientific Reports*, vol. 8, no. 1, article 3045, 2018.
- [10] M. Ling, M. J. Esfahani, H. Akbari, and A. Foroughi, “Effects of residence time and heating rate on gasification of petroleum residue,” *Petroleum Science and Technology*, vol. 34, no. 22, pp. 1837–1840, 2016.
- [11] H. L. Ma and S. B. Tsai, “Design of research on performance of a new iridium coordination compound for the detection of Hg<sup>2+</sup>,” *International Journal of Environmental Research and Public Health*, vol. 14, no. 10, article 1232, 2017.
- [12] L. Y. Mo, W. H. Z. Sun, S. Jiang et al., “Removal of colloidal precipitation plugging with high-power ultrasound,” *Ultrasonics Sonochemistry*, vol. 69, article 105259, 2020.
- [13] D. Gao, Y. Liu, Z. Guo et al., “A study on optimization of CBM water drainage by well-test deconvolution in the early development stage,” *Water*, vol. 10, no. 7, p. 929, 2018.
- [14] S. B. Tsai and H. Ma, “A research on preparation and application of the monolithic catalyst with interconnecting pore structure,” *Scientific Reports*, vol. 8, no. 1, article 16605, 2018.
- [15] J. Xie and H. Ma, “Application of improved APO algorithm in vulnerability assessment and reconstruction of microgrid,” *IOP Conference Series: Earth and Environmental Science*, vol. 108, no. 5, article 052109, 2018.
- [16] A. Fujin, Y. Xiuzhao, and H. Ruochi, “Research into the super-absorbent polymers on agricultural water,” *Water Management*, vol. 245, article 106513, 2021.
- [17] Y. Shen, B. Biondi, and R. Clapp, “Q-model building using one-way wave-equation migration Q analysis — part 2: 3D field-data test,” *Geophysics*, vol. 83, no. 2, pp. S111–S126, 2018.
- [18] R. D. Hume, L. Berry, S. Reichelt et al., “An engineered human adipose/collagen model for in vitro breast cancer cell migration studies,” *Tissue Engineering Part A*, vol. 24, no. 17-18, pp. 1309–1319, 2018.
- [19] M. Patel, S. Chaudhary, and S. Garg, “Improved pre-copy algorithm using statistical prediction and compression model for efficient live memory migration,” *International Journal of High Performance Computing and Networking*, vol. 11, no. 1, pp. 55–65, 2018.
- [20] R. W. Allen, J. S. Collier, A. G. Stewart et al., “The role of arc migration in the development of the lesser Antilles: a new tectonic model for the Cenozoic evolution of the eastern Caribbean,” *Geology*, vol. 47, no. 9, pp. 891–895, 2019.
- [21] D. Xu and H. Ma, “Degradation of rhodamine B in water by ultrasound-assisted TiO<sub>2</sub> photocatalysis,” *Journal of Cleaner Production*, vol. 313, article 127758, 2021.
- [22] J. Hey, Y. Chung, A. Sethuraman et al., “Phylogeny estimation by integration over isolation with migration models,” *Molecular Biology and Evolution*, vol. 35, no. 11, pp. 2805–2818, 2018.
- [23] S. E. Vollset, E. Goren, C. W. Yuan et al., “Fertility, mortality, migration, and population scenarios for 195 countries and territories from 2017 to 2100: a forecasting analysis for the Global Burden of Disease Study,” *The Lancet*, vol. 396, no. 10258, pp. 1285–1306, 2020.
- [24] S. Zhang, A. Lorenzo, C. Zhou, Y. Cui, B. Gonçalves, and M. Angel Gómez, “Performance profiles and opposition interaction during game-play in elite basketball: evidences from National Basketball Association,” *International Journal of Performance Analysis in Sport*, vol. 19, no. 1, pp. 28–48, 2019.

## Research Article

# Distilling the Knowledge of Multiscale Densely Connected Deep Networks in Mechanical Intelligent Diagnosis

Xiaochuan Wang <sup>1</sup>, Aiguo Chen <sup>1</sup>, Liang Zhang,<sup>1,2</sup> Yi Gu <sup>1</sup>, Mang Xu,<sup>1</sup>  
and Haoyuan Yan<sup>1</sup>

<sup>1</sup>School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi, Jiangsu 214122, China

<sup>2</sup>School of Biotechnology, Jiangnan University, Wuxi, Jiangsu 214122, China

Correspondence should be addressed to Aiguo Chen; [agchen@jiangnan.edu.cn](mailto:agchen@jiangnan.edu.cn)

Received 24 May 2021; Revised 20 June 2021; Accepted 22 June 2021; Published 7 July 2021

Academic Editor: Shan Zhong

Copyright © 2021 Xiaochuan Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

At present, deep neural network (DNN) technology is often used in intelligent diagnosis research. However, the huge amount of calculation of DNN makes it difficult to apply in industrial practice. In this paper, an advanced multiscale dense connection deep network MSDC-NET is designed. A well-designed multiscale parallel branch module is used in the network. This module can greatly improve the acceptance domain of MSDC-NET, so as to learn useful information from input samples more effectively. Based on the inspiration of Densely Connected Convolutional Networks, MSDC-NET designed a similar dense connection technology, so that the model will not have the problem of gradient vanishing because of the deep network. The experimental data of MSDC-NET on MFPT, SEU, and Pu datasets show that our method has higher performance than other latest technologies. At the same time, we carried out knowledge distillation based on the high-precision classification level of MSDC-NET, which makes the diagnosis ability and robustness of the lightweight CNN model improve significantly.

## 1. Introduction

The mechanical equipment in modern industry often needs to work in the complex environment of high temperature, fatigue, and heavy load for a long time, which may cause incalculable production accidents and economic property losses. At present, most enterprises use the way of manual supervision, which consumes a lot of human and material resources. Since mechanical failures are often not so obvious, only technicians with professional knowledge can be sure to judge mechanical failures. As a mature unsupervised system, intelligent diagnosis can solve this problem for related enterprises. The traditional intelligent diagnosis method mainly uses a machine learning algorithm to classify the feature data from the sensor signal. For example, Zhang et al. used support vector machine (SVM) to identify the working state of bearing [1]; Bane-

rjee and Das used SVM fusion multisensor signal to detect motor fault [2]; Bugarbee and Trendafilova used the nearest neighbor classifier to identify faults [3]; Keskes et al. studied the method of feature extraction using stationary wavelet packet transform, and support vector machine (SVM) is applied to rotor fault classification. [4]; Mustafa et al. proposed that the spectrum analysis method be used in feature extraction and SVM be used to detect mechanical faults. [5]. However, the signal obtained by sensor needs complex feature extraction, which is difficult to achieve instantaneity, and depends on professional knowledge and relevant experience.

With the rapid development of deep learning (DL) technology in many fields, researchers have begun to use some DL-based technologies to achieve intelligent diagnosis, such as Multilayer Perceptron (MLP) [6], Deep Belief-Net (DBN) [7], Convolutional Neural Net (CNN) [8],

Autoencoder-Net (AE) [9], and Recurrent Neural Network (RNN) [10]. The MLP model is a very simple perceptron model. The knowledge it can learn is very scarce, and it has been rarely used at present. As a probabilistic generation model relative to the traditional discriminant model, DBN is composed of multiple restricted Boltzmann machine layers; Shao et al. proposed to use the optimized DBN model to diagnose rolling bearings [11]; Tamilselvan and Wang proposed a DBN model for engine fault diagnosis [12]. The improvement of the CNN model is mainly in two directions. One is to use different input types, such as two-dimensional images [13], infrared thermal images [14], vibration spectrum images [15], and time-frequency images [16]. Another direction is to improve the structure of CNN model, such as combining the CNN with Generative Adversarial Network (GAN [17]) to generate more new labeled samples [18, 19]. Janssens et al. proposed a CNN model to classify the faults of rotating machinery [20]. Recently, some researchers use transfer learning [21, 22] combined with the CNN to obtain the prior knowledge of the original dataset [23–25]. The AE model is composed of multiple automatic encoders, and there are two common extended models, such as denoising AE (DAE) [26] and compression AE (CAE) [27]. The advantage of DAE is that it can learn useful information from damaged data, and CAE learns more stable feature representations through penalty items. The latest AE model also combines with the GAN network to generate labeled samples [28–30] and also embeds the semisupervised learning method into the VAE model [31, 32]. The Recurrent Neural Network has no advantage in classification, and it is more commonly used in mechanical life prediction.

So as to evaluate the effectiveness and fairness of the model, our experiment is carried out in the benchmark code base proposed by Zhao et al. [33]. They conducted a comprehensive benchmark study on the realization of intelligent diagnosis based on the DL model. The study evaluated the intelligent diagnosis performance of nine commonly used models on nine publicly available datasets. The nine models include three encoder models, such as AE, DAE, and CAE, including the MLP model and CNN model and two advanced deep classification networks AlexNet [34], ResNet18 [35], and finally include an LSTM [36] model. The results of these studies show the significant advantages of DL in intelligent diagnosis, especially the models that have outstanding performance in other classification problems such as ResNet18, which also have good results when transplanted to intelligent diagnosis. However, the scale of these deep classification networks is very large, and the computing power of the equipment is relatively high. How to improve the accuracy of model intelligent diagnosis in different accuracy requirements and equipment environment is a worthy research direction.

The main target of our method is to use a new and powerful deep classification network as the powerful model to guide the lightweight model to realize mechanical intelligent diagnosis. Therefore, our main contribution can be outlined in two aspects:

- (1) We designed a new multiscale densely connected deep network MSDC-NET, and the experimental results of

five comparison algorithms on three types of datasets show that the model has higher superiority

- (2) We innovatively use the knowledge distillation technology in the mechanical intelligent diagnosis subject, which makes the classification accuracy of the lightweight CNN model significantly improve and provides an effective solution to achieve the highest level of diagnosis in the case of limited computing power

The experiments in this paper are carried out on the open MFPT bearing dataset (<https://mfpt.org/fault-data-sets/>), Pu bearing dataset (<https://mb.uni-paderborn.de/kat/forschung/dacenter/bearing-dacenter/>), and SEU transmission dataset (<https://github.com/cathysiyu/Mechanical-datasets>).

The rest of the paper is arranged as follows: Section 2 retrospectively the related work of deep classification network and knowledge distillation technology, and then in Section 3, we introduce our MSDC-NET model, knowledge distillation process, and its algorithm in detail. After that, the experimental results obtained on three common datasets are discussed and analyzed in Section 4. Finally, the article is summarized in the Section 5.

## 2. Related Work

*2.1. Deep Convolution Network.* The CNN, which was proposed in 1997, is a classification network for processing labeled data. It can transfer, extract, and learn information through convolution and pool operations. AlexNet won the championship in the 2012 ImageNet competition, and the ResNet proposed in 2016 has surpassed human classification accuracy for the first time.

The CNN model is generally composed of a convolution layer, pooling layer (maximum pooling layer, average pooling layer), and fully connected layer. We often use the convolution layer and maximum pooling layer with a step size of 2 to learn features and finally classify features through full connection layer.

The convolution layer operation is equivalent to the multiplication of input  $x$  and convolution kernel  $w$ , which is then represented by the activation function mapping:

$$h_k^l = \phi(w_k^l \cdot x + b_k^l), \quad (1)$$

where  $\phi$  is the convolution operator,  $h_k^l$  is the output of convolution operation,  $k$  is the  $k$ -th convolution kernel,  $l$  is the  $l$ -th layer, and  $w_k^l$  and  $b_k^l$  are the weight and bias.

The purpose of maximum pooling layer is to learn the most essential knowledge in the local acceptance domain, focus on texture information, and reduce the bias of estimation mean caused by convolution layer parameter error. The purpose of average pooling layer is to consider the characteristics of step size, focus on the background information, and reduce the estimation variance.

As the number of layers of CNN tends to be deeper, the problem of model degradation may arise, and jump connection technology is often used to avoid gradient disappearance. Meanwhile, the multiscale technology widely used in

CV field can also be transplanted into the CNN to improve the receptive field of network model.

**2.2. Distilling the Knowledge.** In the industrial application, in addition to the requirement that the model should have as high a prediction level as possible, it is also expected that the expenditure of the model should be as small as possible, so that the deployment needs the least computing resources (computing power, storage space) and has a lower delay. Knowledge distillation [37] is a solution to this problem. First, we need a powerful and effective pretraining model (T-Model), then train a lightweight model (S-Model) from scratch and pass on the knowledge contained in a T-Model to an S-Model during the training process.

The above process of knowledge transfer only requires that the “SoftMax” output distribution of S-Model and T-Model under a given input be fully closed.

$$q_i = \frac{\exp(z_i/\text{TEMP})}{\sum_j \exp(z_j/\text{TEMP})}, \quad (2)$$

where  $q_i$  is the output distribution of S-Model,  $z_i$  is the calculated probability for each class in S-Model, and TEMP is a parameter usually set to 1. When TEMP becomes larger, we get a softer output. For cross-entropy loss, the gradient for a logit of S-Model is as follows:

$$\frac{\partial C}{\partial z_i} = \frac{1}{\text{TEMP}}(q_i - p_i) = \frac{1}{\text{TEMP}} \left( \frac{e^{z_i/\text{TEMP}}}{\sum_j e^{z_j/\text{TEMP}}} - \frac{e^{c_i/\text{TEMP}}}{\sum_j e^{c_j/\text{TEMP}}} \right), \quad (3)$$

where  $p_i$  is the output distribution of the T-Model,  $c_i$  is the calculated probability for each class in the T-Model, and when  $x$  tends to 0, it is equivalent to infinitesimal with  $e^x - 1$ . Easy to know, when TEMP is sufficiently large, there are

$$\frac{\partial C}{\partial z_i} \approx \frac{1}{\text{TEMP}} \left( \frac{1 + (z_i/\text{TEMP})}{N + \sum_j z_j/\text{TEMP}} - \frac{1 + (c_i/\text{TEMP})}{N + \sum_j c_j/\text{TEMP}} \right). \quad (4)$$

Suppose that every sample has Logits with zero mean, that is  $\sum_j z_j = \sum_j c_j = 0$ , then:

$$\frac{\partial C}{\partial z_i} \approx \frac{1}{N\text{TEMP}^2}(z_i - c_i). \quad (5)$$

### 3. Our Proposed Method

**3.1. The Composition of the Blocks.** To ensure the performance of the model, we have carefully designed the neural network modules shown in Figures 1–3, through the stacking of these modules constituting the final MSDC-NET.

Go\_Down cell is proposed to achieve a similar dense connection effect, which can maximize the retention of the original lore of the previous data, and realize the superposition of the previous and subsequent feature maps in the channel dimension. The “times” represents the multiple relationship of the size of the previous and subsequent feature maps. After

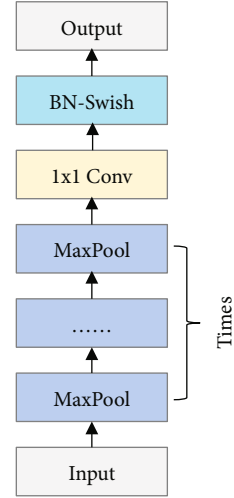


FIGURE 1: The construction of a Go\_Down cell.

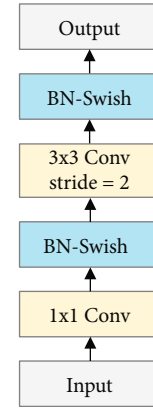


FIGURE 2: The construction of a DS cell.

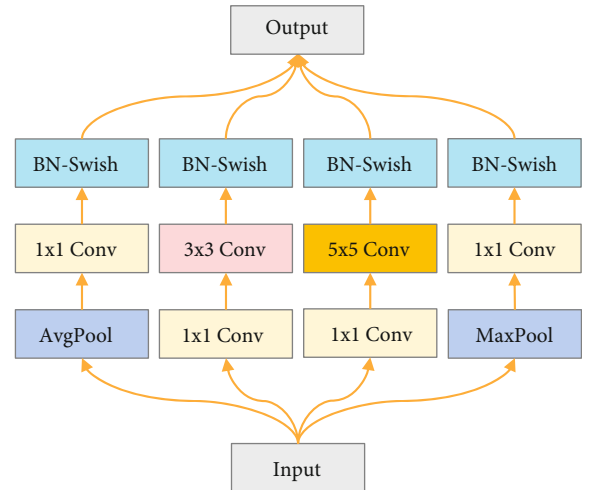


FIGURE 3: The construction of an MS\_Block cell.

“times” max pooling, the dimension of the feature map will be reduced to 1/times of the original size.

DS cell is proposed for downsampling of the feature map, which is implemented by convolution operation (the

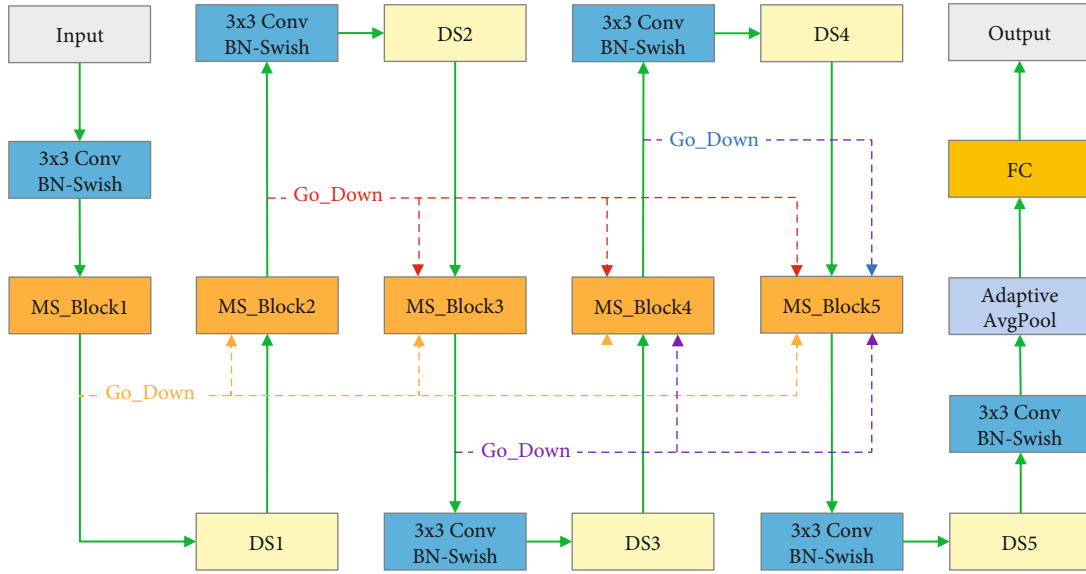


FIGURE 4: The structure of the MSDC-NET.

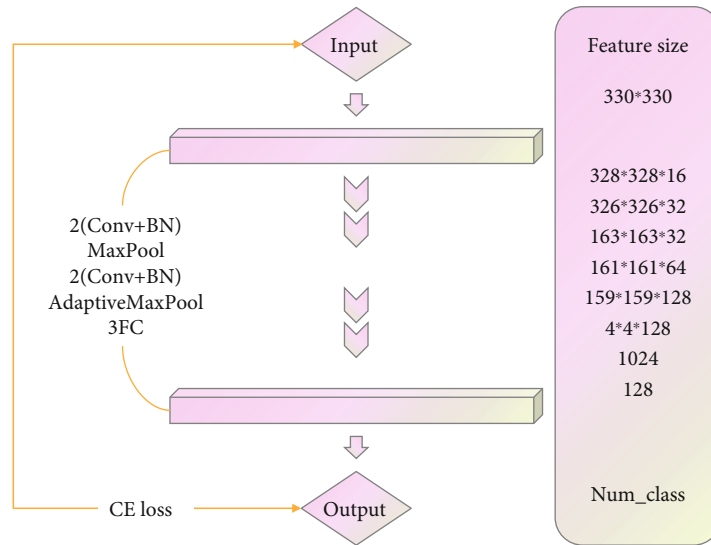


FIGURE 5: The structure of the CNN.

parameters are as follows: stride = 2 and kernel\_size = 3 × 3). The dimension of the input data will be decreased by half, and the number of channels will be doubled.

MS\_Block cell draws on the idea of inception cell [38], in which 3 × 3 and 5 × 5 convolution kernels are used for multiscale learning feature map, and average pooling and maximum pooling are used to extract useful information from multiple angles. After the multibranch network structure has learned the input data, a complete feature map of each branch superimposed and fused in the channel dimension will be obtained, and then, this feature map will be placed in the MSDC network to continue learning.

3.2. The Structure of MSDC-NET. For the purpose of get a high-precision Teacher-Model, we integrated the most prev-

alent multiscale and dense connection technology to advance a split-new deep classification network MSDC-NET. It is made up of our carefully designed stack of various modules. Taking the input data size AB as an example, five MS\_Block cells are set in the model to realize multiscale and multiangle learning sample data information. At the same time, five DS cells are set for downsampling, and the size of the top feature map is AC. In practical applications, we can adjust the network model reasonably according to the input data size. When the input data is large, more stacking blocks can be set, and on the contrary, the number of layers can be set fewer.

However, only through the one-way propagation of this order often fails to achieve a good classification effect, because the deep network will forget and lose a lot of previous

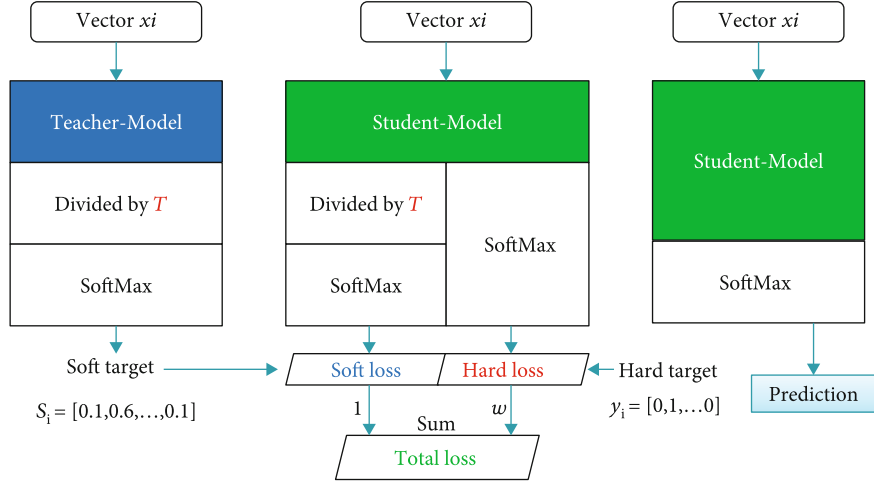


FIGURE 6: Flow chart of knowledge distillation technology.

Input: Sample X, Sample category, Test sample  $\tilde{x}$

Step1: The training process of the MSDC- NET model:

1) Processing sample input data:

Unify the sample data X into the same segmentation method, input type and input size, and finally we get the input as (x, y).

2) Train the MSDC-NET model:

Train MSDC-NET through the processed training data, and the objective function (6) as:

$$\min loss = -\sum_{i=1}^n (l \log(\tilde{l}) + (1-l) \log(1-\tilde{l}))$$

Step2: The training process of the CNN model:

1) Processing sample input data:

2) Train the CNN model:

Train CNN through the processed training data, and the objective function (10) as

$$\min loss = \omega * loss\_hard + loss\_soft$$

Step3: Fault classification

1) Output of MSDC-NET (or CNN):

Input the test data sample  $\tilde{x}$  into the pre-trained model (MSDC-NET or CNN) to get the output  $\tilde{y}$ .

2) Category determination

After obtaining the output  $\tilde{y}$  of the model (MSDC-NET or CNN), use Formula (7) to predict the category of the test sample:

$$label(\tilde{x}) = \arg \max(\tilde{y})$$

Output: Predict the category of input x

ALGORITHM 1: Process of MSDC-NET (or CNN) for fault classification

knowledge and information. We need a unique cascade operation to fuse the same resolution features of different stages in the channel dimension, so we designed a technology similar to dense connection. The implementation of this technology relies on Go\_Down cell, through which we save the output feature map of each MS\_Block cell. The feature maps obtained by Go\_Down of different "times" can be merged with the small feature map of matching size later, which is able to speed up the convergence of the network model, while improving the robustness of the model and avoiding problems such as gradient disappearance.

The structure of the MSDC-NET model we proposed is shown in Figure 4. The main road after the data sample enters the model is the green one-way line. A network training procedure using reasonable and effective loss function is crucial; the output  $\tilde{l}$  is obtained from the top-level feature

graph through Adaptive AvgPool and FC layers. If the class label of data sample is  $l$ , the loss function of model training is as follows:

$$loss(Cross\ entropy) = -\sum_{i=1}^n (l \log(\tilde{l}) + (1-l) \log(1-\tilde{l})). \quad (6)$$

Since the quantity of layers of the DNN is relatively deep, the useful information contained in the first data entering the network and the first feature map may be forgotten and lost. We borrowed the idea of DenseNet [39] in the MSDC-NET model and, after the foregoing features of FIG different sizes corresponding to multiples by scaling Go\_Down cell extract, then merged them with the feature maps of the back of the

TABLE 1: Fault classification of PU dataset.

Code	Fault mode: description
K001	Health state: launched 50 hours in advance
K002	Health state: launched 19 hours in advance
K003	Health state: launched 1 hours in advance
K004	Health state: launched 5 hours in advance
K005	Health state: launched 10 hours in advance
K006	Health state: launched 16 hours in advance
KA01	Outer ring: man-made damage caused by EDM (Level 1)
KA03	Outer ring: man-made damage caused by electric engraver (Level 2)
KA05	Outer ring: man-made damage caused by electric engraver (Level 1)
KA06	Outer ring: man-made damage caused by electric engraver (Level 2)
KA07	Outer ring: man-made damage caused by drilling (Level 1)
KA08	Outer ring: man-made damage caused by drilling (Level 2)
KA09	Outer ring: man-made damage caused by drilling (Level 2)
KI01	Inner ring: man-made damage caused by EDM (Level 1)
KI03	Inner ring: man-made damage caused by electric engraver (Level 1)
KI05	Inner ring: man-made damage caused by electric engraver (Level 1)
KI07	Inner ring: man-made damage caused by electric engraver (Level 2)
KI08	Inner ring: man-made damage caused by electric engraver (Level 2)
KA04	Outer ring: damage caused by fatigue and pitting (single point and single damage and Level 1)
KA15	Outer ring: damage caused by plastic deform and indentation (single point and single damage and Level 1)
KA16	Outer ring: damage caused by fatigue and pitting (single point and repetitive damage and Level 2)
KA22	Outer ring: damage caused by fatigue and pitting (single point and single damage and Level1)
KA30	Outer ring: damage caused by plastic deform and indentation (distributed and repetitive damage and Level 1)
KB23	Outer ring, inner ring: damage caused by fatigue and pitting (single point and multiple damage and Level 2)
KB24	Outer ring, inner ring: damage caused by fatigue and pitting (distributed and multiple damage and Level 3)
KB27	Outer ring, inner ring: damage caused by fatigue and pitting (distributed and multiple damage and Level 1)
KI04	Inner ring: damage caused by fatigue and pitting (single point + single damage + Level 1)
KI14	Inner ring: damage caused by fatigue and pitting (single point + multiple damage + Level 1)
KI16	Inner ring: damage caused by fatigue and pitting (single point + single damage + Level 3)
KI17	Inner ring: damage caused by fatigue and pitting (single point + repetitive damage + Level 1)
KI18	Inner ring: damage caused by fatigue and pitting (single point + single damage + Level 2)
KI21	Inner ring: damage caused by fatigue and pitting (single point + single damage + Level 1)

TABLE 2: Fault classification of MFPT dataset.

Category	Description	Category	Description
Health state	270 lb, 25 Hz, 97656 sps,6 s	/	/
Outer ring 1	25 lb, 25 Hz, 48828 sps, 3 s	Inner ring 1	25 lb, 25 Hz, 48828 sps, 3 s
Outer ring 2	50 lb, 25 Hz, 48828 sps, 3 s	Inner ring 2	50 lb, 25 Hz, 48828 sps, 3 s
Outer ring 3	100 lb, 25 Hz, 48828 sps, 3 s	Inner ring 3	100 lb, 25 Hz, 48828 sps, 3 s
Outer ring 4	150 lb, 25 Hz, 48828 sps, 3 s	Inner ring 4	150 lb, 25 Hz, 48828 sps, 3 s
Outer ring 5	200 lb, 25 Hz, 48828 sps, 3 s	Inner ring 5	200 lb, 25 Hz, 48828 sps, 3 s
Outer ring 6	250 lb, 25 Hz, 48828 sps, 3 s	Inner ring 6	250 lb, 25 Hz, 48828 sps, 3 s
Outer ring 7	300 lb, 25 Hz, 48828 sps, 3 s	Inner ring 7	300 lb, 25 Hz, 48828 sps, 3 s



TABLE 3: Fault classification of SEU dataset.

Sort	Fault mode (RS-LC:20 Hz-0 V)	Sort	Fault mode (RS-LC:3 Hz-2 V)
1	Health gear	11	Health gear
2	Health bearing	12	Health bearing
3	Chipped tooth	13	Chipped tooth
4	Inner ring	14	Inner ring
5	Missing tooth	15	Missing tooth
6	Outer ring	16	Outer ring
7	Root fault	17	Root fault
8	Inner + outer rings	18	Inner + outer rings
9	Surface fault	19	Surface fault
10	Rolling element	20	Rolling element

network. And we let them influence the training process of the network to varying degrees according to the set weights.

Finally, the category of input data is forecasted by the following formula:

$$label(x) = \arg \max (\tilde{l}). \quad (7)$$

**3.3. CNN after Knowledge Distillation.** We use the simple CNN model shown in Figure 5. With the input by way of multiple stacked Conv layers and Pool layers, it enters the classifier to obtain the final output.

The model contains only 4 convolutional layers in total, with a very simple structure and very low requirements on the computing power and storage of the deployed equipment, so it is suitable to run in the industrial environment.

**3.4. Algorithm Details.** In Section 2, we describe the process of knowledge distillation in detail. After knowledge distillation, the CNN model can obtain  $loss\_soft$  according to formula (5). (In the experiment, we set  $T$  as 10 according to the relevant literature.)

$$loss\_soft = \frac{1}{NT^2} (z_i - v_i). \quad (8)$$

According to the previous formula (6) of MSDC-NET, we can get  $loss\_hard$  as follows:

$$loss\_hard = - \sum_{i=1}^n \left( l \log (\tilde{l}) + (1-l) \log (1-\tilde{l}) \right). \quad (9)$$

Finally, the total loss function of knowledge distillation process is (where  $\omega$  is the weight of  $loss\_hard$ )

$$loss = \omega * loss\_hard + loss\_soft. \quad (10)$$

As shown in Figure 6, after the Teacher-Model is divided by the temperature parameter  $T$ , the soft label is obtained through the ‘‘SoftMax’’ transformation, and the value in the label is between 0 and 1. The larger the temperature parameter  $T$ , the softer the distribution. On the contrary, it is easy

to introduce unnecessary noise and amplify the probability of misclassification. We need to make sure that the correct predicted contribution is made in the Teacher-Model, so set  $T$  to a larger value. The real label in the sample is converted to a one-hot vector and used as a hard label. The total loss is the weighted sum of the cross entropy of the soft label and the predicted value and the cross entropy of the hard label and the predicted value. If the value of  $\omega$  is smaller, it indicates that more attention is paid to the contribution of Teacher-Model; on the contrary, let the Student-Model pay more attention to the identification of difficult samples. Therefore, the higher the classification accuracy of the teacher model, the more conducive to student model learning samples. Our specific algorithm is reflected in Algorithm 1.

**Algorithm 1:** Process of MSDC-NET (or CNN) for fault classification  
Input: Sample X, Sample category, Test sample  $\tilde{x}$

Step1: The training process of the MSDC-NET model:

1) Processing sample input data:

Unify the sample data X into the same segmentation method, input type and input size, and finally we get the input as (x, y).

2) Train the MSDC-NET model:

Train MSDC-NET through the processed training data, and the objective function (6) as:

$$\min loss = - \sum_{i=1}^n (l \log (\tilde{l}) + (1-l) \log (1-\tilde{l}))$$

Step2: The training process of the CNN model:

1) Processing sample input data:

2) Train the CNN model:

Train CNN through the processed training data, and the objective function (10) as

$$\min loss = \omega * loss\_hard + loss\_soft$$

Step3: Fault classification

1) Output of MSDC-NET (or CNN):

Input the test data sample  $\tilde{x}$  into the pre-trained model (MSDC-NET or CNN) to get the output  $\tilde{y}$ .

2) Category determination

After obtaining the output  $\tilde{y}$  of the model (MSDC-NET or CNN), use Formula (7) to predict the category of the test sample:

$$label(\tilde{x}) = \arg \max (\tilde{y})$$

Output: Predict the category of input x

## 4. Experiment and Discussion

### 4.1. Datasets

**4.1.1. PU Bearing Dataset.** There are 21 categories, including 6 undamaged bearings, 12 damaged bearings, and 14 actual

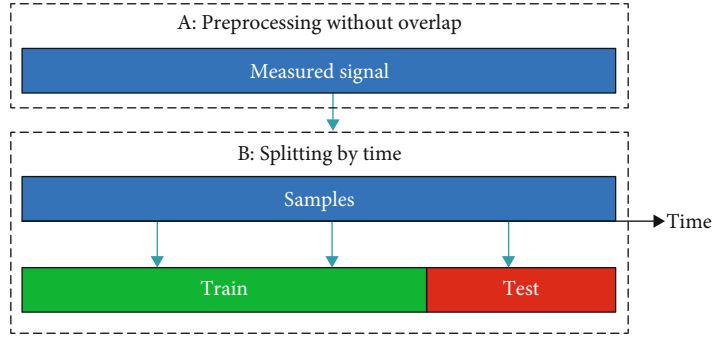


FIGURE 7: Data split according to time sequences.

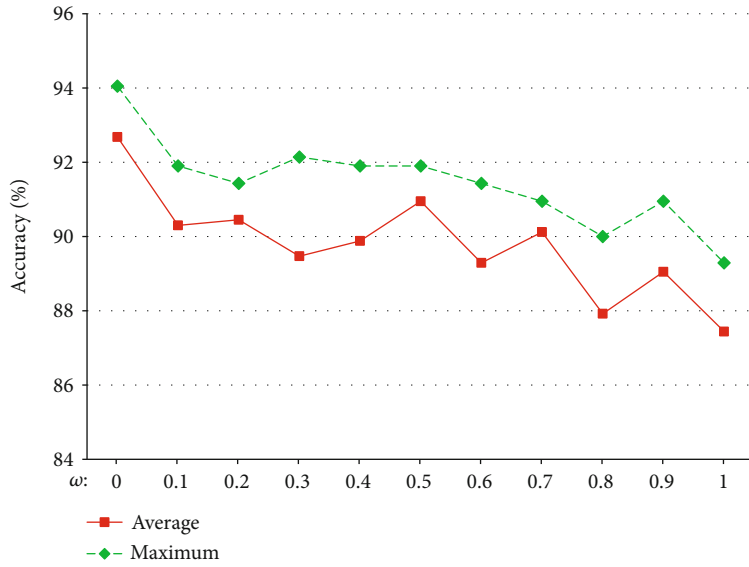


FIGURE 8: The influence of different  $\omega$  on the model.

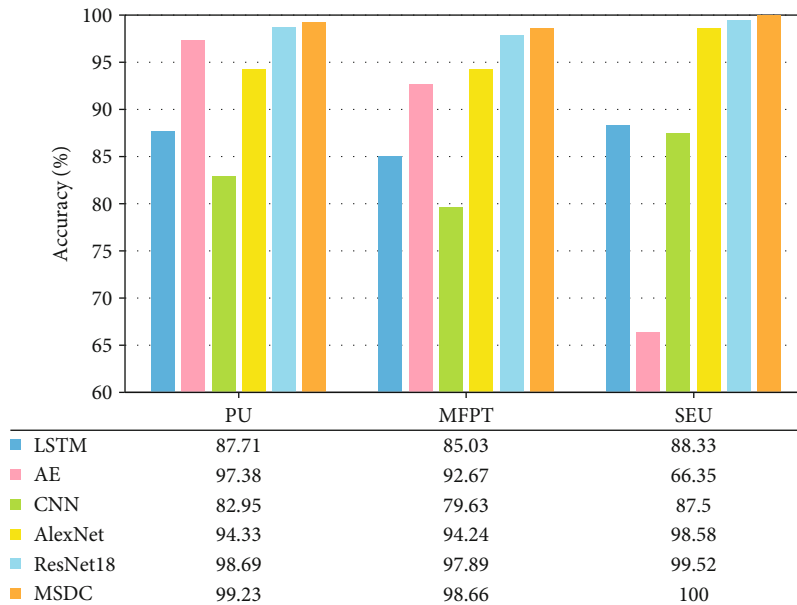


FIGURE 9: Maximum prediction accuracy graph of MSDC-NET and five contrast algorithms on three datasets.

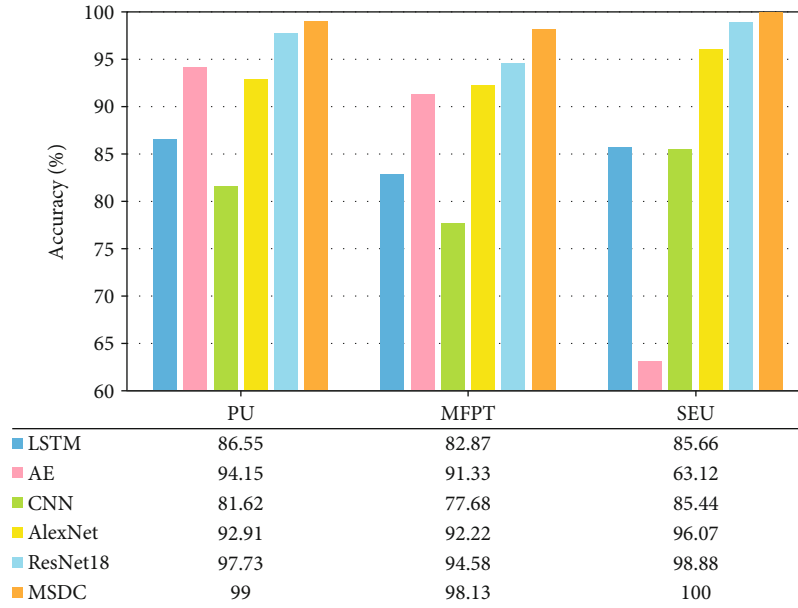


FIGURE 10: Average prediction accuracy graph of MSDC-NET and five contrast algorithms on three datasets.

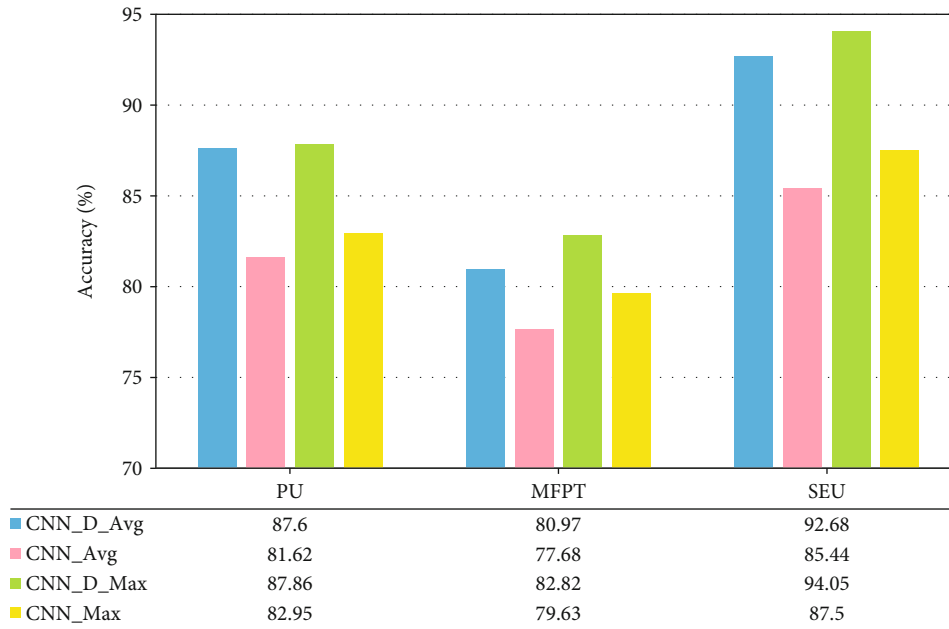


FIGURE 11: Map of average and maximum classification accuracy of original CNN and knowledge distilled CNN\_D on three datasets.

damaged bearings caused by an accelerated life test. The details are shown in Table 1.

**4.1.2. MFPT Bearing Dataset.** There are 15 categories, according to different loads, one healthy bearing and fourteen faulty bearings. The fault classification is displayed in Table 2. The four values in the Description are load, input shaft speed, sampling rate, and duration.

**4.1.3. SEU Gearbox Dataset.** There are 20 categories, including inner ring, outer ring, and rolling element in different

states. In each file, we use the second line of eight vibration signals. Details are shown in Table 3.

**4.2. Data Split.** There are two common data segmentation methods in mechanical intelligent diagnosis. One is the stochastic segmentation strategy, and the other is the time sort segmentation of datasets. The random segmentation strategy may cause data overlap between two sets, resulting in the risk of test leakage, and the final test accuracy cannot be used as a basis for fair evaluation of model performance. Industrial data are generally continuous, which may contain

time-dependent knowledge information. Therefore, in order to evaluate the scientific and retain the useful information in the dataset as much as possible, we adopt the strategy of dividing the dataset according to the time sort (Figure 7). The last 20% of the time series data is used for testing, and the rest is used for training.

**4.3. Input Types.** In mechanical intelligent diagnosis, signal processing methods are often used to map sample data to other domains. These methods include converted into time domain (TD), converted into frequency domain (FD), converted into time-frequency domain (TFD), converted into slice image (SI), and converted into wavelet domain (WD). The benchmark study of Zhao et al. [33] showed that FD and TFD can achieve higher accuracy, so we finally used TFD to complete the experiment.

$$x_i^{\text{STFT}} = \text{STFT}(x_i), \quad i = 1, 2, \dots, N. \quad (11)$$

Inside  $\text{STFT}(\cdot)$  is to convert  $x_i$  into TFD.

TFD is to perform short-time Fourier transform (STFT) on each specimen, and the Hanning window length is 64. Since the CNN model requires a larger 2D size to extract features, the final signal size is adjusted to  $330 \times 330$ .

**4.4. Training Details.** During model training, we used the Adam optimizer. Some parameters of model training are as follows: learning rate = 0.001 and batch size = 8 (due to the device limitation). Each model is trained and tested alternately during training, and a total of 100 epochs are experienced. All comparison models and MSDC-NET are trained and fairly compared under the open source code framework proposed by Zhao et al. [33], and the data enhancement method provided in the code framework is adopted. All work is performed on the device equipped with E5-2640 V4 @ CPU and 12G GeForce RTX 2080 Ti GPU. The device system is Ubuntu 16.04, the running environment is Python 3.6, and the Pytorch1.4 library is used.

We have done the optimization experiment on the SEU gearbox dataset for the value of  $\omega$  between 0 and 1. It is able to deduce from Figure 8 that as  $\omega$  increases, the performance of the CNN model will tend to the performance without knowledge distillation; classification accuracy showed a downward trend, which also shows the effectiveness of knowledge distillation technology from the side, so we set  $\omega$  to 0 in the subsequent comparative experiments.

**4.5. Evaluating Indexes.** In our research process, the comprehensive accuracy (Acc) is used to fairly assess the capability of the algorithm. With the purpose of avoiding the impact of the fluctuation of DNN model in training, we repeat each experiment 10 times. Finally, we take the average Acc and the maximum Acc as the assessment indexes, the average Acc is the average of 10 times of Acc, and the maximum Acc is the maximum of 10 times of Acc.

**4.6. MSDC-NET Experimental Results.** For each algorithm, we completed 10 tests on the corresponding dataset in the same environment. The achievement of the highest predicted Acc is recorded in Figure 9, and then, the achievement of the

average predicted Acc of 10 times is recorded in Figure 10. From the overall view of these two figures, our algorithm reached the first-rate capability. The LSTM algorithm does not perform well on classification problems, and the average test Acc is between 85% and 90%.

Compared with those on other datasets, the capability of the AE network on the SEU dataset is significantly not the same, which shows that the AE network has higher requirements for datasets. The classification ability of the simple CNN network is relatively poor without knowledge distillation, ranging from 75 to 85. The test Acc of DNN is relatively high, the test Acc of AlexNet is between 95 and 100, the test Acc of ResNet18 is between 98 and 100, and the test Acc of our MSDC-NET is between 98 and 100. From the comparison chart of average test Acc, our method is 1.27%, 3.55%, and 1.12% higher than ResNet18, which is the best algorithm in comparison. In the maximum test Acc diagram, our method is 0.54%, 0.77%, and 0.48% higher than ResNet18. Meanwhile, it is a surprise that our MSDC-NET has achieved 100% results on SEU gearbox dataset for 10 times, which reflects the powerful classification ability and robustness of our algorithm and will not produce serious over fitting phenomenon due to the deep network layers.

**4.7. CNN Experimental Results after Knowledge Distillation.** We distill the knowledge of MSDC-NET according to the algorithm in Figure 6, so that the classification Acc of the CNN model is significantly improved without any change in the internal structure. As can be seen in Figure 11, contrasted with the original CNN, the average test Acc of CNN\_D after knowledge distillation is 5.98%, 3.29%, and 7.24% higher on the three datasets. The maximum test Acc is 4.91%, 3.19%, and 6.55% higher, respectively. The average test Acc of CNN\_D even exceeds the maximum test Acc of the original CNN. CNN\_D has the most obvious improvement on the SEU gearbox dataset, which is comparable to the deep neural classification network AlexNet shown in Figure 10. Due to the problems of MFPT dataset itself, the effect of knowledge distillation is not obvious, but it has also been improved.

## 5. Conclusion

In this paper, we created a new MSDC-NET, which is derived from several typical deep convolutional neural network structures, combined with the most advanced multiscale and jump connection technologies. Experimental results on several datasets show that our MSDC-NET has a more prominent classification level than other latest research. At the same time, we found that knowledge of distillation technology in the diagnosis of intelligent mechanical issue is effective, and the classification accuracy of small CNN model after knowledge distillation is significantly improved. Next, we will further study the knowledge distillation technology, improve the diagnostic accuracy of the CNN model after knowledge distillation, and verify the portability and reliability of the related technology in the direction of mechanical intelligence prediction.

## Data Availability

The labeled dataset used to support the findings of this study is available from the corresponding author upon request.

## Conflicts of Interest

The authors declare no conflicts of interest.

## Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 61772241 and Grant U20A20228.

## References

- [1] X. L. Zhang, B. J. Wang, and X. F. Chen, "Intelligent fault diagnosis of roller bearings with multivariable ensemble-based incremental support vector machine," *Knowledge-Based Systems*, vol. 89, pp. 56–85, 2015.
- [2] T. P. Banerjee and S. Das, "Multi-sensor data fusion using support vector machine for motor fault detection," *Information Sciences*, vol. 217, pp. 96–107, 2012.
- [3] H. al-Bugharbee and I. Trendafilova, "A fault diagnosis methodology for rolling element bearings based on advanced signal pretreatment and autoregressive modelling," *Journal of Sound and Vibration*, vol. 369, pp. 246–265, 2016.
- [4] H. Keskes, A. Braham, and Z. Lachiri, "Broken rotor bar diagnosis in induction machines through stationary wavelet packet transform and multiclass wavelet SVM," *Electric Power Systems Research*, vol. 97, pp. 151–157, 2013.
- [5] M. O. Mustafa, D. Varagnolo, G. Nikolakopoulos, and T. Gustafsson, "Detecting broken rotor bars in induction motors with model-based support vector classifiers," *Control Engineering Practice*, vol. 52, pp. 15–23, 2016.
- [6] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, *Learning Internal Representations by Error Propagation*, California Univ San Diego La Jolla Inst for Cognitive Science, 1985.
- [7] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
- [8] Y. LeCun and Y. Bengio, "Convolutional networks for images, speech, and time series," *The handbook of brain theory and neural networks*, vol. 3361, p. 1995, 1995.
- [9] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," 2014, <http://arxiv.org/abs/1312.6114>.
- [10] W. Zaremba, I. Sutskever, and O. Vinyals, "Recurrent neural network regularization," 2014, <http://arxiv.org/abs/1409.2329>.
- [11] H. D. Shao, H. K. Jiang, X. Zhang, and M. G. Niu, "Rolling bearing fault diagnosis using an optimization deep belief network," *Measurement Science and Technology*, vol. 26, no. 11, p. 115002, 2015.
- [12] P. Tamilselvan and P. F. Wang, "Failure diagnosis using deep belief learning based health state classification," *Reliability Engineering and System Safety*, vol. 115, pp. 124–135, 2013.
- [13] J. Zhang, Y. Sun, L. Guo, H. Gao, X. Hong, and H. Song, "A new bearing fault diagnosis method based on modified convolutional neural networks," *Chinese Journal of Aeronautics*, vol. 33, no. 2, pp. 439–447, 2020.
- [14] Y. Li, X. Du, F. Wan, X. Wang, and H. Yu, "Rotating machinery fault diagnosis based on convolutional neural network and infrared thermal imaging," *Chinese Journal of Aeronautics*, vol. 33, no. 2, pp. 427–438, 2020.
- [15] A. Youcef Khodja, N. Guersi, M. N. Saadi, and N. Boutasseta, "Rolling element bearing fault diagnosis for rotating machinery using vibration spectrum imaging and convolutional neural networks," *The International Journal of Advanced Manufacturing Technology*, vol. 106, no. 5-6, pp. 1737–1751, 2020.
- [16] Y. Zhang, K. Xing, R. Bai, D. Sun, and Z. Meng, "An enhanced convolutional neural network for bearing fault diagnosis based on time-frequency image," *Measurement*, vol. 157, p. 107667, 2020.
- [17] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza et al., "Generative adversarial networks," *Advances in Neural Information Processing Systems*, vol. 3, pp. 2672–2680, 2014.
- [18] D. B. Verstraete, E. L. Droguett, V. Meruane, M. Modarres, and A. Ferrada, "Deep semi-supervised generative adversarial fault diagnostics of rolling element bearings," *Structural Health Monitoring*, vol. 19, no. 2, pp. 390–411, 2020.
- [19] W. Zhang, X. Li, X.-D. Jia, H. Ma, Z. Luo, and X. Li, "Machinery fault diagnosis with imbalanced data using deep generative adversarial networks," *Measurement*, vol. 152, article 107377, 2020.
- [20] O. Janssens, V. Slavkovicj, B. Vervisch et al., "Convolutional neural network based fault detection for rotating machinery," *Journal of Sound and Vibration*, vol. 377, pp. 331–345, 2016.
- [21] Y. Jiang, X. Gu, D. Wu et al., "A novel negative-transfer-resistant fuzzy clustering model with a shared cross-domain transfer latent space and its application to brain CT image segmentation," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 18, no. 1, pp. 1–52, 2020.
- [22] Y. Jiang, K. Zhao, K. Xia et al., "A novel distributed multi-task fuzzy clustering algorithm for automatic MR brain image segmentation," *Journal of Medical Systems*, vol. 43, no. 5, 2019.
- [23] G. Xu, M. Liu, Z. Jiang, W. Shen, and C. Huang, "Online fault diagnosis method based on transfer convolutional neural networks," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 2, pp. 509–520, 2020.
- [24] W. Mao, L. Ding, S. Tian, and X. Liang, "Online detection for bearing incipient fault based on deep transfer learning," *Measurement*, vol. 152, article 107278, 2020.
- [25] Z. Chen, K. Gryllias, and W. Li, "Intelligent fault diagnosis for rotary machinery using transferable convolutional neural network," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 1, pp. 339–349, 2020.
- [26] Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proceedings of the 25th international conference on Machine learning - ICML '08*, pp. 1096–1103, Helsinki, Finland, 2008.
- [27] M. Ranzato, C. Poultney, S. Chopra, and Y. L. Cun, "Efficient learning of sparse representations with an energy-based model," *Advances in Neural Information Processing Systems*, vol. 19, pp. 1137–1144, 2007.
- [28] X. Xiong, J. Hongkai, X. Li, and M. Niu, "A Wasserstein gradient-penalty generative adversarial network with deep auto-encoder for bearing intelligent fault diagnosis," *Measurement Science and Technology*, vol. 31, no. 4, p. 045006, 2020.

- [29] F. Zhou, S. Yang, H. Fujita, D. Chen, and C. Wen, "Deep learning fault diagnosis method based on global optimization GAN for unbalanced data," *Knowledge-Based Systems*, vol. 187, p. 104837, 2020.
- [30] Q. Guo, Y. Li, Y. Song, D. Wang, and W. Chen, "Intelligent fault diagnosis method based on full 1-D convolutional generative adversarial network," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 3, pp. 2044–2053, 2020.
- [31] N. Jiang, X. Hu, and N. Li, "Graphical temporal semi-supervised deep learning-based principal fault localization in wind turbine systems," *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, vol. 234, no. 9, pp. 985–999, 2020.
- [32] X. Li, J. Li, Y. Qu, and D. He, "Semi-supervised gear fault diagnosis using raw vibration signal based on deep learning," *Chinese Journal of Aeronautics*, vol. 33, no. 2, pp. 418–426, 2020.
- [33] Z. Zhao, T. Li, J. Wu et al., "Deep learning algorithms for rotating machinery intelligent diagnosis: an open source benchmark study," *ISA Transactions*, vol. 107, pp. 224–255, 2020.
- [34] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, pp. 1097–1105, 2012.
- [35] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, Las Vegas, NV, USA, 2016.
- [36] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [37] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," *Computer Science*, vol. 14, no. 7, pp. 38–39, 2015.
- [38] G. Huang, Z. Liu, V. Laurens, and K. Q. Weinberger, "Densely connected convolutional networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9, Honolulu, HI, USA, 2017.
- [39] C. Szegedy, W. Liu, Y. Jia et al., "Going deeper with convolutions," in *2015 IEEE conference on computer vision and pattern recognition (CVPR)*, Boston, MA, USA, 2015.

## Research Article

# Evolutionary Algorithm for Multiobjective Optimization Based on Density Estimation Ranking

Lin Li, Hengfei Wu , Xiujian Hu, and Guanglei Sheng

Department of Electronics and Information Engineering, Bozhou University, Bozhou Anhui, China

Correspondence should be addressed to Hengfei Wu; whf520315000@163.com

Received 9 May 2021; Revised 1 June 2021; Accepted 19 June 2021; Published 6 July 2021

Academic Editor: Shan Zhong

Copyright © 2021 Lin Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In the past few decades, a number of multiobjective evolutionary algorithms (MOEAs) have been proposed in the continue study. As pointed out in some recent studies, the performance of the most existing MOEAs is not promising when solving different shapes of Pareto fronts. To address this issue, this paper proposes an MOEA based on density estimation ranking. The algorithm includes density estimation ranking to shift the reference solution position, calculating the density of candidate solutions and ranking by the estimated density value, to modify the Pareto dominance relation and for handling complicated Pareto front. The result of this ranking can be used as the second selection criterion for environmental selection, and the optimal candidate individual with distribution and diversity information is selected. Experimental results show that the proposed algorithm can solve various types of Pareto fronts, outperformance several state-of-the-art evolutionary algorithms in multiobjective optimization.

## 1. Introduction

Multiobjective optimization problems (MOPs) are common in the real-life, e.g., robotics [1], urban bus transit route network design problem [2], smart grids [3], and electricity selling market [4]. These problems aim to optimize more than two conflicting objectives at the same time, which can be mathematically formulated clearly depict as follows.

$$\begin{aligned} \text{Min} \quad & F(x) = (f_1(x), f_2(x), \dots, f_m(x)) \\ \text{s.t} \quad & x \in X \end{aligned} \quad (1)$$

where  $\mathbf{x}$  denotes the decision vector,  $X \subseteq R^n$  is called the decision space, and  $F(x) \in R^m$  is the objective vector of  $\mathbf{x}$ , which consists of  $m$  objective functions of  $f_i(x)$ ,  $i = 1, 2, \dots, m$ .

However, owing to the conflicting nature of multiple objectives, there is not an algorithm that obtains a single optimal solution that can optimize all objectives. Instead, some solutions can be obtained as a trade-off between different objectives, called the Pareto set (PS). The PS is termed the

Pareto front (PF) in the objective space. To approximate the Pareto optimal set, greatly quantities of multiobjective evolutionary algorithms (MOEAs) have been proposed to solve MOPs in the past several decades. These algorithms can be roughly classified into three categories.

The first category is the dominance-based MOEAs, which keep the nondominated solution and remove the dominated solution in the population [5]. There are mainly two types of dominance-based approaches: the first type of Pareto dominance-based MOEAs, which Pareto dominance-based mechanisms are adopted to distinguish and select candidate solutions. For example, the elitist nondominated sorting genetic algorithm (NSGA-II) [6], the Pareto envelop-based selection algorithm II (PESAI) [7], and the strength Pareto evolutionary algorithm 2 (SPEA2) [8] are representative MOEAs of this type, where all nondominated solutions are firstly identified, and then the other strategy is used to make selections among the nondominated solutions to maintain the population diversity.

The second type of dominance-based approaches, where it improves the selection pressure of the environment by

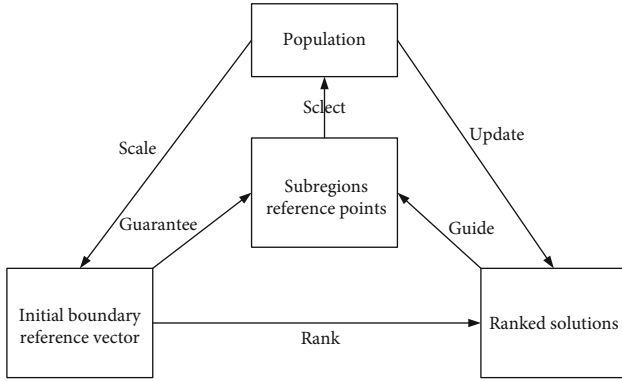


FIGURE 1: The relationships between the four sets in DERE, i.e., the population  $P$ , the initial boundary reference vectors  $R$ , the ranked solutions  $S$ , and the subregions boundary reference point set  $R'$ . First, the initial vectors of scale in  $R$  according to the dimension of the population  $P$  and the extreme solutions is selected by the boundary reference vector and obtain the rank value 0. Then, the ranked solutions  $S$  close the true PF as subregions boundary reference point  $R'$ . Finally, the DER selects subregions optimal solution and rank.

changing the Pareto dominance, thus does not need for other selection strategies. Remarkable MOEAs based on change Pareto dominance have included  $\epsilon$ -MOEA [9], MO\_Ring\_PSO\_SCD [10], and HTL-PSO [11].

The second category is commonly known as the decomposition-based MOEAs, which decomposes a complex MOP into a number of subproblems and optimizes them simultaneously [12]. There are mainly divided MOEAs into two types of decomposition-based approaches. In the first type of decomposition-based approaches, an MOP is divided into a group of single-objective optimization problems (SOPs) [13]. The MOEA/D proposed in [14] is a simple and generic MOEA based on decomposition, which introduces a set of weight vectors to manage and select candidate solutions. [15] proposed a reference vector-guided evolutionary algorithm, which MOPs are divided into a set of SOPs by the reference vector; in this way, the candidate solutions will effectively converge to the optimal solutions of each SOPs without considering the conflict between different SOPs.

In the second type of decomposition-based approaches, an MOP is divided into a set of sub-MOPs or subregions. For instance [16], the during decades proposed include reference point-based nondominated sorting genetic algorithm III (NSGA-III) [17], inverse modeling-based MOEA (IM-MOEA) [18], and the  $k$ -means clustering method divides the population into multiple subpopulations in GSMPSO-MM [19], and learning to-decompose paradigm adaptively sets the decomposition method with the Pareto front [20, 21] that proposed a MOEAD-M2M, which divided MOP into a group of simple subproblems and solving these in a collaborative manner, population diversity will be achieved by this way. The proposed algorithm termed as SPEA/R, which divides objective space into a set of subregions, and individuals in every subregion are driven toward target direction [22].

The third category is known as the indicator-based approaches, where performance indicators of solution qual-

```

Input:  $N$  (population size)
Output:  $P$  (a set of solutions)
1 Initialize:  $P = \text{RandomInitialize}(N)$ ;
2 while termination criterion not fulfilled do
3  $P' = \text{Matingselection}(2, N, P)$ 
4  $O = \text{GA}(P')$ 
5  $P = \text{EnvironmentalSelection}(P \cup O, N)$ 
6 end while
  
```

ALGORITHM 1: Framework of DERE.

ity measurement are adopted as selection criteria in the environmental selection. Representative MOEAs of this type contain  $S$  metric selection evolutionary multiobjective optimization algorithm (SMS-EMOA) [23], generational distance and  $\epsilon$ -dominance based MOEA (GDE-MOEA) [24], indicator-based evolutionary algorithm [25], a dynamic neighborhood MOEA based on hypervolume indicator (DNMOEA/HI) [26], an inverted generational distance (IGD) indicator-based evolutionary algorithm [27], and R2 indicator-based many-objective metaheuristic-II (MOMBI-II) [28], where the environment selection strategies are designed based on a predefined the hypervolume (HV) indicator, the generation distance (GD), binary indicator, and the R2 indicator, respectively. The hype suggested an HV indicator-based MOEA [29], where it uses Monte Carlo simulation to approximate the exact HV values. The AR-MOEA is based on the enhanced inverted GD (IGD-NS) indicator [30], where a reference point adaptation method is to adjust a set of reference points based on the indicator calculation of candidate solutions.

There are a few other algorithms that are not included in the above three main categories. For example, a decision variable analysis MOEA (MOEA/DVA) has been proposed based on decomposition and differential evolution (DE) [31]. The MOEA/D-CMA suggested in [32] in which it uses DE and covariance matrix in the MOEA based on decomposition, and DREA has been proposed in the [33], which is based on diversity ranking. More recently, a new algorithm has been suggested based on fuzzy and decomposition for multiobjective optimization (MOEA-MCD) [34]. An enhanced two-archive algorithm (Two\_Arch2) has been suggested based on different selection principles (indicator and dominance) in [35], which design a new Lp-norm maintenance for diversity and convergence. A one-by-one selection strategy has been provided in [36], where main approach is in the environment selection strategy to selection offspring individuals one by one, which is based on a computationally efficient convergence indicator to increase the selection pressure towards the Pareto optimal front.

These MOEAs are suggested to solve MOPs in the literature, and the experimental results verify that many MOEAs can well balance the convergence and diversity of population. And yet, it have been pointed out in some recent studies that the most existing MOEAs for the optimization problems show performance that can strongly depend on PF shape [37, 38]. In other words, some MOEAs are merely solving a type of PFs of MOP and show versatility that is not good when these MOEA



```

Input:  $P$  (population)
Output:  $P'$  (parents for variation)
1 Define the rank of all solutions by density estimation ranking;
2 Calculate the density estimation DE value of all solutions;
3 for  $i = 1$  to  $|P|$  do
4 Randomly select  $p$  and  $q$  from  $P$ ;
5 if  $DE_p > DE_q$  then
6  $P' = P' \cup q$ ;
7 else if  $DE_p < DE_q$  then
8  $P' = P' \cup p$ ;
9 else
10  $P' = P' \cup \text{random}(p, q)$ ;
11 end if
12 end for

```

ALGORITHM 2: Mating selection.

optimizing other types PFs of MOP. This is due to the most MOEAs that are very sensitive to PF shapes.

For the above issues, this paper has been suggested a density estimation ranking- (DER-) based evolutionary algorithm for multiobjective optimization, called DERE. The main contributions of this work included as follows.

- (1) The current work has much of the study primary selection criterion (i.e., modifying the Pareto dominance approach) in Pareto-based algorithms to solved multiobjective, and this paper enhances the Pareto-based algorithm performance by the other selection criterion (modifying the diversity maintenance mechanism). This paper has been suggested a DER approach in order to enhance Pareto-based algorithm suitable for multiobjective optimization. The DER simultaneously includes the convergence and distribution information of individuals to enhance MOEA performance
- (2) The basic approach of DER is simple, considering the contribution of individuals to population convergence in sparse area, and the DER shifts poor convergence individuals to density crowded area. In this way, these individuals with poor convergence performance will be given a high density value, which is easier to be eliminated in the evolution process, so improving the efficiency of the algorithm. In addition, the calculated density value is used directly for ranking, and the calculation cost is negligible

The rest of this paper is organized as follows. In Section 2, the details of the proposed algorithm DERE are described. The empirical results of DERE compared with several classical MOEAs are presented in Section 3. Finally, conclusion and future work are given in Section 4.

## 2. The Proposed DERE Preparation

*2.1. The Framework of DERE.* The proposed DERE has similar framework as NASA-II, except that DER is adopted

as the second select criterion and the boundary reference vectors to manage diversity and convergence for various types of the Pareto fronts. In general, there are four main solution sets maintained in DERE, i.e., the population  $P$ , the initial boundary reference vectors  $R$ , the ranked solutions  $S$ , and the subregions boundary reference point set  $R'$ . To be specific, the population  $P$  contains the candidate solutions, guaranteeing uniform distribution of the candidate solutions in  $P$  by the initial boundary reference vector set  $R$  and the ranked solutions  $S$  as final output reflects the PF and guides the boundary reference points adaptation come into being, and the subregions boundary reference point set  $R'$  is used in the DER-based selection in the population  $P$ , where the relationships between the four solutions are described in Figure 1.

As presented in Algorithm 1, the main framework of DERE consists of the following steps. Firstly, an initial population  $P$  of size  $N$  is randomly generated. Then, in the main loop, the mating selection is carried out to construct the  $P'$  (mating pool) based on the ranking results, where mating selection will be detail described in Algorithm 2. The  $O$  (offspring population) has been generated by genetic algorithm operations. The parents and their offspring solutions are obtained by the environment selection. Finally, the best  $N$  solutions are selected by the environment selection, and the new population comes into being.

*2.2. Offspring Creation.* In the suggested DERE, the widely used genetic operators [39], i.e., the polynomial mutation [40] and the SBX (simulated binary crossover) [41], are introduced to combine with the binary tournament selection [42], which constructs the mating selection method to generate the offspring population. Thus, initial population  $P$  includes  $N$  individuals, and a number of  $N/2$  couple of parents are randomly generated, i.e., each of the  $N$  individuals has same probability to participate in the reproduction procedure. Algorithm 2 details the binary tournament selection and density estimation in the main loop. This is possible partly thanks to boundary reference vector strategy (it will be explained in detail later), which can capture the boundary

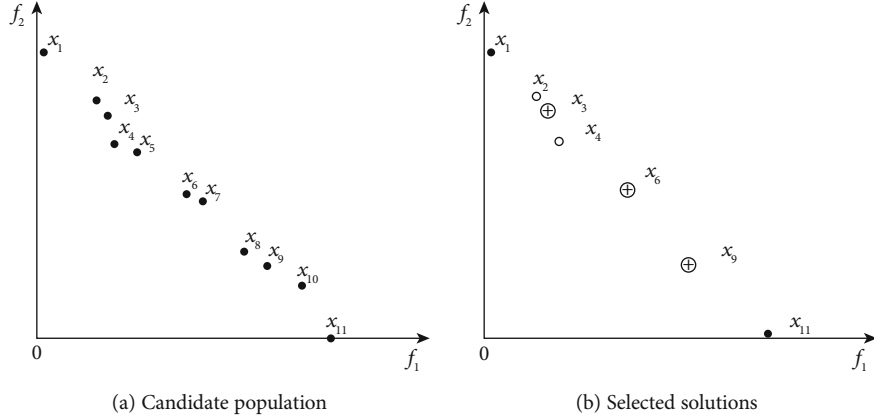


FIGURE 2: An illustrative example to show the advantage of DER over SDE.

of the PF to manage the diversity effectively ( $m$  boundary reference vector search  $m$  solutions).

**2.3. Density Estimating.** In a population, an individual density represents the degree of crowding in the region where the individual is located. Density estimation plays a fundamental role in the select Pareto solution set. There are various density estimation strategies in the MOEAs. For instance, the SPEA2 uses a nearest neighbor density estimation strategy to calculate the density of an individual [8]. NSGA-II gives an expression in the density of an individual by calculating crowding distance [6]. Most grid-based MOEAs, for example, the dynamic multiobjective evolutionary algorithm [43] and PESAI, calculate the number of individuals in the hyperbox to reflect the density of an individual; yet, some scholars proposed to use a set of hyperboxes to describe the density of an individual based on the degree of individual crowding [44, 45].

Despite the various strategies of density estimation, this all measures the similarity of degree between individuals in a population. Generally, lower density individuals are preferable when two individuals are nondominant individuals in the population. DE contributes greatly to the Pareto solution set in the MOEA-based Pareto dominates. Therefore, this paper introduces enhance shift-based density estimating (SDE) [46] to maintain the diversity and convergence of population.

Simply speaking, if there is an individual performing better than individual  $p$  for an objective, this objective will be moved to the same position of  $p$  in the population; otherwise, its position will remain unchanged. Formally, the density  $D(p, P)$  of a candidate solution  $p$  in the population  $P$  can be expressed as follows:

$$D(p, P) = D\left(\text{dist}\left(p, q'_1\right) + \text{dist}\left(p, q'_2\right) + \dots + \text{dist}\left(p, q'_{N-1}\right)\right), \quad (2)$$

where  $\text{dist}(p, q'_i)$  denotes the calculation of the Euclidean distance between individual  $p$  and selected solution  $q'_i$ .  $N$  denotes the size of population  $P$ , and  $q'_i$  is the shift version of an indi-

**Input:**  $P = \{x_1, x_2, \dots, x_N\}$ : Population.

**Output:**  $R(x \in P)$ : Ranking results.

```

1 for  $i = 1$  to  $m$  do
2 Find the  $i$ -th extreme solution  $e_i$ 
3  $R(e_i) = 0$  //extreme solutions obtain the rank 0
4 end for
5  $O = \{e_1, e_2, \dots, e_m\}$  // the set of ranked solutions
6  $U = P - O$  // the set of unranked solutions
7 for all  $x \in U$  do
   //initialize the distance to ranked solutions
8  $DE(x) = \min(\text{dist}(x, e'_1) + \dots + \text{dist}(x, e'_m))$ 
   //  $e'_i$  Shifted by  $e_i$ .
9 end for.
10  $R(x) = DE(x)$  // current rank value
11 while  $|U| > 0$  do
   // find the candidate solution close to PF in region  $s$ 
12  $s = \{x | x \in U, \min(\text{dist}(x, O'_s))\}$ 
   //  $O'_s$  has a subregions boundary solution  $O_s$  shifted
13  $R(s) = DE(s)$ 
14  $O = O \cup \{s\}, U = U - \{s\}$ 
15 end while

```

ALGORITHM 3: Density estimation ranking.

vidual  $q_i$  ( $q_i \in P$  and  $q_i \neq p$ ), which is defined as follows:

$$q'_{i(j)} = \begin{cases} q_{i(j)} & \text{otherwise} \\ p_{(j)} & \text{if } q_{i(j)} < p_{i(j)} \end{cases}, j \in (1, 2, 3, \dots, m), \quad (3)$$

where  $p_{(j)}$ ,  $q_{i(j)}$ , and  $q'_{i(j)}$  denote the  $j$ th objective value of individual  $p$ ,  $q_i$ , and  $q'_i$  respectively, and  $m$  denotes the number of objectives.

The following sections will describe the two main components of suggested DERE, i.e., the density estimation ranking and environment selection.

**2.4. Ranking the Solutions.** When a great number of objectives are involved, the proportion of nondominant individuals in the population becomes quite large. In extreme cases, all individuals in the population may become

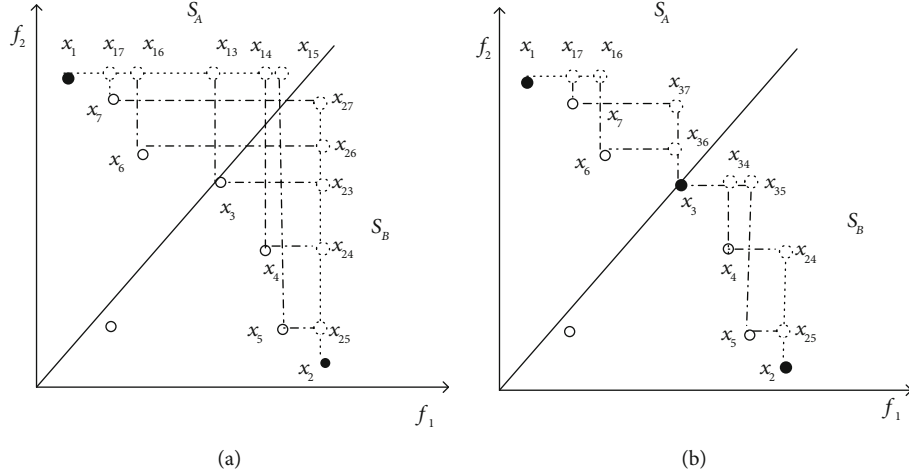


FIGURE 3: Illustration of the first two iterations of the ranking approach.

nondominated with each other. In this case, the individual density will play a dominant or even unique role in the selection process of the algorithm to distinguish them. It can be clearly known that only the individual with both good diversity and convergence has a low crowding value, when using the density estimation method calculation of an individual density. Certainly, either individual with poor diversity or poor convergence has some neighbors. Individual has a high crowding when an individual has with both poor diversity and poor convergence. Therefore, we want to select the individual with the smallest density value by ranking.

Based on the density estimation strategy, the DER is proposed in this subsection. Supposing a population  $P$  contains  $N$  solutions, the task of DER is to distribute a rank value to each candidate solution in  $P$ , using the calculation density value of the individual as the rank value and describes its quality in terms of both diversity and convergence. In this study, hoping to get a smaller rank value, the best solutions accept the rank 0. Therefore, when applying density estimation to evolutionary algorithms, candidate solutions with smaller rank values are more apt to survive in the mating selection and the environmental selection.

Compared to the SDE, DER adopts more comprehensive information to select a solution set. Consider an example as shown in Figure 2, where  $x_1-x_{11}$  is a candidate population. In the case that five out of the eleven candidate solutions are to be selected for next generation, SDE to select solutions is  $\{x_1, x_2, x_3, x_4, x_{11}\}$ . The DER chooses evenly distributed solutions that are  $\{x_1, x_3, x_6, x_9, x_{11}\}$ , which are the best candidate solutions in terms of diversity and convergence. The reason for this phenomenon is that SDE selected the enough solutions in the one region. However, during the selection process by DER category, solutions divide a complex PF into a number of subregions, then selecting the optimal solution in every subregion.

DER works like a selection operator which selects unranked solutions iteratively. In its main loop, there are two sets  $O$  and  $U$ , which are composed of ranked solutions and unranked individuals, respectively. Initially,  $O$  is the set

```

Input:  $P$  (combined population),  $N_V$  (reference vector),
 $N$  (population size)
Output:  $Q$  (population for next generation)
// Non-dominated sorting.
1  $Front = \text{NondominatedSort}(P)$ ;
2  $k \leftarrow$  minimum number satisfies  $|\cup_{i=1}^k Front_i| \geq N$ ;
3  $Q \leftarrow \cup_{i=1}^{k-1} Front_i$ ;
4 // * solutions obtained from last front * l
5  $p \leftarrow \arg \min_{p \in Front_k} N_V$ ;
6 if  $p + Q \geq N$ 
7  $Q = Q \cup p$ 
8 else continue select solutions
9 while  $length < (N - Q - p)$ 
10  $Q' = \arg \max (DER(Front_k))$ ;
11 end
12  $Q = Q \cup Q' \cup p$ 
13 end

```

ALGORITHM 4: Environmental selection.

of extreme solutions  $e_1, e_2, \dots, e_m$ , and the extreme solutions are selected by the boundary reference vector. Capturing the boundary of the PF can be beneficial for the next step of the selection optimal solutions. This part is not the main contribution of this paper, and thus the method proposed in [16] is employed here, and the extreme solutions in  $O$ . These are the most boundary solutions of PF, facilitating the DER method to rank the population by PF boundary. However,  $U$  includes all remaining solutions:

$$O = \{e_1, e_2, \dots, e_m\}, U = P - O. \quad (4)$$

After the initialization, we select a solution  $s$  with the smallest individual density value from  $U$  according to the density estimation method. The  $s$  is selected processing

which is defined as follows:

$$s = \arg \min_{x \in U} D(x, e_O), \quad (5)$$

where  $e_O$  denotes the extreme solutions by boundary reference vectors, and  $D(x, e_O)$  denotes calculate density of an individual  $x$  in the population. This step is designed to maintain the diversity. Since  $s$  is close to PF, selecting the solutions has high quality in terms of both convergence and diversity. The estimated individual density value as the ranking value is moved to  $O$ . Then, the  $s$  is divided in the entire region into two subregions, and selecting optimal solution  $s$  in every subregions is moved to  $O$ . These operations are repeated until  $U$  is empty.

All the details of DER have been discussed in above. Algorithm 3 provides the pseudocode of DER. The result of ranking  $x$  is denoted by  $R(x)$ . Lines 1 to 4 use the boundary vector to select the extreme solutions and obtain the rank 0. In lines 5 to 9,  $O$  and  $U$  are initializing, and the minimum density value to the rank solutions after the shifted position of each solution in  $U$  is initializing. Then, loop in lines 11 to 16 iteratively ranks the remaining solutions. Specifically, in constant iteration, the promise solution  $s$  is first found in line 12. The density value of the promise solution is its rank value in line 13. Finally, in line 14, we move  $s$  from  $U$  to  $O$ . Since the key individual is inserted every time the loop ranks, the density of unranked individuals will gradually decrease; so, previously ranked solutions always get higher rank value than those ranked in the subsequent iterations.

To clearly understanding the ranking of process, Figure 3 depicts an illustration of its first two iterations. The circle in the figure represents the solutions in the entire population, where filled circle (i.e.,  $x_1$  and  $x_2$ ) represents the extreme solutions selected by the boundary vector, and the dotted circle indicates the position after the extreme solution is shifted when calculating the individual density.  $x_{13}, x_{14}, x_{15}, x_{16}, x_{17}, x_{23}, x_{24}, x_{25}, x_{26}$ , and  $x_{27}$  denote the position after shifting when calculating the individual density, where  $x_{13}$  and  $x_{23}$  represent the position after  $x_1$  and  $x_2$  shift when calculating the density of individual  $x_3$ . Firstly, calculating the density of unranked individual  $p$  using the extreme solutions selected by the boundary vector, where extreme solution position is shifted by the density estimation approach and then selected the individual with the smallest density value as the promise solution  $s$ . Then, the individual  $x_3$  with the smallest density value is selected as  $s$  by comparing the density values of all individuals. Meanwhile, it gets rank value that is the density value. Secondly, bipartition the objective space into two subregions  $S_A$  and  $S_B$  by  $x_3$ . Then,  $x_1$  and  $x_3$  are the boundary extreme solutions of subregion  $S_A$ , and  $x_2$  and  $x_3$  are the boundary extreme solutions of subregion  $S_B$ ; in the second iteration to select key individual in each subregion, the key individual with the smallest density value is the promising solution, and it gets the rank and is moved to  $O$ . After these iterations, obtaining the rank solutions is  $x_1, x_2, x_3, x_4$ , and  $x_7$ .

In the above example, it is easy to understand that the selection pressure is usually greater in the previous stage, but the selection pressure gradually decreases with more and more

TABLE 1: Pareto fronts of the test instances.

Problem	Pareto front	
DTLZ1	Linear	
DTLZ2-4	Concave	Regular
WFG4-9	Concave	
DTLZ5-6	Mostly degenerate	
DTLZ7	Disconnected	
IDTLZ1-2	Inverted	
WFG1	Sharp tails	Irregular
WFG2	Disconnected	
WFG3	Mostly degenerate	

TABLE 2: Setting of the population size.

$M$	Division	Population size
2	(15, 0)	100
3	(13, 0)	105

solutions that are selected and ranked. Thus, previously ranked solutions have a high quality, because they have survived in the fierce selection competition. And they will be assigned a larger rank values compared to those ranked in the later stage.

*2.5. Environment Selection.* The DERE is similar to the most existing MOEAs, which adopt elite strategies to make environmental selections for the combined population of each generation's parent and offspring candidate solutions.

The procedure of the DER-based environmental selection is given in Algorithm 4. The main loops are as follows: Firstly, before the DER to selecting the candidate solution, making the combined population is first sorted by adopting the high efficiency nondominated rank. Then, population in all candidate solutions from nondominated fronts in the first  $k-1$  was selected directly for the next generation (line 3 in Algorithm 4). Secondly, the corresponding solution is selected by the boundary vector to enter the next generation (line 5 in Algorithm 4). Finally, because  $p + Q < N$ , continuing to select  $N - (p + Q)$  candidate solutions from the population by DER (line 11 and 13 in Algorithm 3). The DER is used to select candidate solutions in the  $k$ th front  $\text{Front}_k$ , where  $k$  defined the minimum number it satisfies  $|\bigcup_{i=1}^k \text{Front}_i| \geq N$ . Each candidate solutions in the  $\text{Front}_k$  and its contribution to  $\text{Front}_k$  on DER are calculating the Euclidean distance. Then, we will delete the solution with small contribution for efficiency, recalculate the contribution of the remaining solution in the  $\text{Front}_k$ , and repeating this process until the number of remaining solutions in  $\bigcup_{i=1}^k \text{Front}_i$  reaches  $N$ .

It is noteworthy that although the selection process in the most Pareto-based dominance MOEAs are guided by the density estimation value, the motivation of employing DER in DERE is very different. In DERE, we transfer position of each individual by SDE mechanism, then rank of each individual by density estimation value. The most existing

TABLE 3: The statistical results (mean and standard deviation) of the HV values obtained by DERA, dMOPSO, MOEA/D, NSGA-II, and SPEA2 on WFG1-WFG9, DTLZ1-DTLZ7, IDTLZ1, and IDTLZ2. The best results are highlighted in blue.

Problem	M	DERA	dMOPSO [48]	MOEA/D [14]	NSGA-II [6]	SPEA2 [8]
WFG1	2	4.1423e-1 (9.71e-2)	1.6041e-1 (3.03e-3)	2.7890e-1 (7.97e-2)	5.0250e-1 (1.03e-1)	4.9385e-1 (1.02e-1)
	3	8.9305e-1 (1.13e-1)	2.9871e-1 (2.93e-3)	9.0517e-1 (1.09e-2)	9.2926e-1 (3.80e-3)	9.4560e-1 (1.30e-3)
WFG2	2	6.3225e-1 (6.05e-4)	5.6806e-1 (8.70e-3)	6.1346e-1 (8.99e-3)	6.3241e-1 (4.18e-4)	6.3222e-1 (6.65e-4)
	3	9.3247e-1 (8.91e-4)	8.0585e-1 (1.19e-2)	9.0801e-1 (5.71e-3)	9.1971e-1 (2.77e-3)	9.2999e-1 (1.17e-3)
WFG3	2	5.8083e-1 (8.02e-4)	5.2468e-1 (5.97e-3)	5.7593e-1 (2.62e-3)	5.7951e-1 (8.29e-4)	5.8056e-1 (9.63e-4)
	3	3.8185e-1 (8.14e-3)	2.7504e-1 (1.75e-2)	3.6930e-1 (2.31e-3)	4.0187e-1 (2.14e-3)	3.8915e-1 (3.98e-3)
DTLZ5	2	3.4783e-1 (1.24e-5)	3.1026e-1 (4.69e-3)	3.4718e-1 (1.38e-5)	3.4650e-1 (1.67e-4)	3.4726e-1 (8.17e-5)
	3	2.0046e-1 (1.54e-5)	1.5808e-1 (7.31e-3)	1.8330e-1 (5.12e-6)	1.9932e-1 (1.81e-4)	1.9976e-1 (1.17e-4)
DTLZ6	2	3.4785e-1 (3.63e-5)	3.4721e-1 (2.68e-6)	3.4722e-1 (1.72e-7)	3.4633e-1 (2.18e-4)	3.4757e-1 (3.71e-5)
	3	2.0048e-1 (9.92e-6)	1.8337e-1 (2.00e-5)	1.8329e-1 (1.48e-6)	1.9963e-1 (1.01e-4)	2.0019e-1 (3.22e-5)
DTLZ7	2	2.4289e-1 (2.62e-5)	2.3606e-1 (1.20e-2)	2.0530e-1 (3.24e-2)	2.4270e-1 (5.73e-5)	2.4285e-1 (2.14e-5)
	3	2.7996e-1 (4.24e-4)	2.4905e-1 (3.92e-3)	2.3227e-1 (1.54e-2)	2.4651e-1 (6.79e-3)	2.5649e-1 (5.88e-3)
IDTLZ1	2	5.4950e-1 (1.09e-1)	2.9142e-4 (1.60e-3)	5.5003e-1 (7.22e-2)	5.7722e-1 (4.33e-3)	5.7570e-1 (9.00e-3)
	3	1.4791e-1 (7.84e-2)	1.3710e-3 (4.06e-3)	1.5541e-1 (4.60e-2)	1.8006e-1 (4.72e-2)	1.8747e-1 (4.63e-2)
IDTLZ2	2	8.1770e-1 (2.49e-4)	8.0144e-1 (1.99e-3)	8.1367e-1 (1.54e-4)	8.1801e-1 (1.63e-4)	8.1874e-1 (9.23e-5)
	3	5.3194e-1 (1.42e-3)	4.6224e-1 (7.18e-3)	5.2589e-1 (2.09e-3)	5.1588e-1 (3.75e-3)	5.3109e-1 (1.86e-3)
		+/- / ≈	0/16/0	3/13/0	5/9/2	6/8/2
WFG4	2	3.4443e-1 (2.34e-3)	2.9411e-1 (4.47e-3)	3.2648e-1 (1.03e-2)	3.4410e-1 (1.47e-3)	3.4405e-1 (2.02e-3)
	3	5.5201e-1 (2.11e-3)	4.4191e-1 (6.91e-3)	5.4790e-1 (1.49e-3)	5.2566e-1 (3.75e-3)	5.4476e-1 (2.16e-3)
WFG5	2	3.1056e-1 (2.39e-3)	3.0388e-1 (2.40e-3)	3.0304e-1 (3.16e-3)	3.1177e-1 (1.90e-3)	3.1159e-1 (2.13e-3)
	3	5.1896e-1 (2.60e-3)	4.3939e-1 (1.37e-2)	5.0433e-1 (4.23e-3)	4.9598e-1 (3.02e-3)	5.1513e-1 (1.63e-3)
WFG6	2	3.0527e-1 (1.44e-2)	2.8863e-1 (8.60e-3)	2.7990e-1 (2.04e-2)	2.9957e-1 (1.02e-2)	3.0383e-1 (1.26e-2)
	3	5.1240e-1 (7.66e-3)	4.1425e-1 (1.94e-2)	4.9717e-1 (1.66e-2)	4.7815e-1 (1.05e-2)	5.0280e-1 (1.25e-2)
WFG7	2	3.4562e-1 (9.44e-4)	2.8086e-1 (6.90e-3)	3.1820e-1 (9.98e-3)	3.4375e-1 (7.48e-4)	3.4468e-1 (7.96e-4)
	3	5.5991e-1 (9.81e-4)	3.9178e-1 (1.04e-2)	5.3659e-1 (3.16e-3)	5.2144e-1 (5.39e-3)	5.4446e-1 (2.40e-3)
WFG8	2	2.8365e-1 (4.66e-3)	2.1843e-1 (8.76e-3)	2.7073e-1 (9.60e-3)	2.8439e-1 (3.02e-3)	2.8474e-1 (3.75e-3)
	3	4.6976e-1 (2.28e-3)	3.0292e-1 (1.19e-2)	4.6876e-1 (2.00e-3)	4.4510e-1 (3.02e-3)	4.6050e-1 (2.60e-3)
WFG9	2	3.3514e-1 (3.41e-3)	3.1497e-1 (6.03e-3)	3.0283e-1 (1.87e-2)	3.3647e-1 (3.92e-3)	3.3555e-1 (3.61e-3)
	3	5.3300e-1 (3.74e-3)	4.6095e-1 (9.46e-3)	5.0223e-1 (2.46e-2)	5.0392e-1 (6.77e-3)	5.1332e-1 (2.12e-2)
DTLZ1	2	5.7960e-1 (2.60e-3)	4.5778e-2 (9.58e-2)	5.7494e-1 (5.18e-3)	5.7954e-1 (1.42e-3)	5.7948e-1 (1.89e-3)
	3	8.4016e-1 (1.90e-3)	2.9001e-2 (8.94e-2)	7.3526e-1 (2.52e-1)	7.6825e-1 (1.75e-1)	7.4473e-1 (2.45e-1)
DTLZ2	2	3.4782e-1 (1.78e-5)	3.0985e-1 (4.28e-3)	3.4718e-1 (1.34e-5)	3.4654e-1 (2.08e-4)	3.4721e-1 (1.14e-4)
	3	5.6487e-1 (4.93e-4)	3.9780e-1 (1.69e-2)	5.6302e-1 (1.76e-6)	5.3343e-1 (3.68e-3)	5.5662e-1 (1.04e-3)
DTLZ3	2	1.3894e-1 (1.41e-1)	1.4711e-3 (7.10e-3)	1.2997e-1 (1.27e-1)	2.3663e-1 (1.25e-1)	2.3811e-1 (1.22e-1)
	3	4.0889e-1 (1.90e-1)	4.1748e-3 (2.13e-2)	0.0000e+0 (0.00e-0)	0.0000e+0 (0.00e-0)	0.0000e+0 (0.00e-0)
DTLZ4	2	3.1352e-1 (8.88e-2)	2.9847e-1 (2.05e-2)	2.9592e-1 (1.04e-1)	2.8696e-1 (1.10e-1)	3.0454e-1 (9.72e-2)
	3	5.3570e-1 (7.50e-2)	4.0682e-1 (3.40e-2)	4.9600e-1 (1.21e-1)	5.2067e-1 (8.13e-2)	4.9832e-1 (1.14e-1)
		+/- / ≈	1/19/0	1/18/1	3/15/2	3/15/2

' + , ' - ' and ' ≈ ' indicates that the result is significantly better, significantly worse, and statistically similar to that obtained by DERA, respectively.

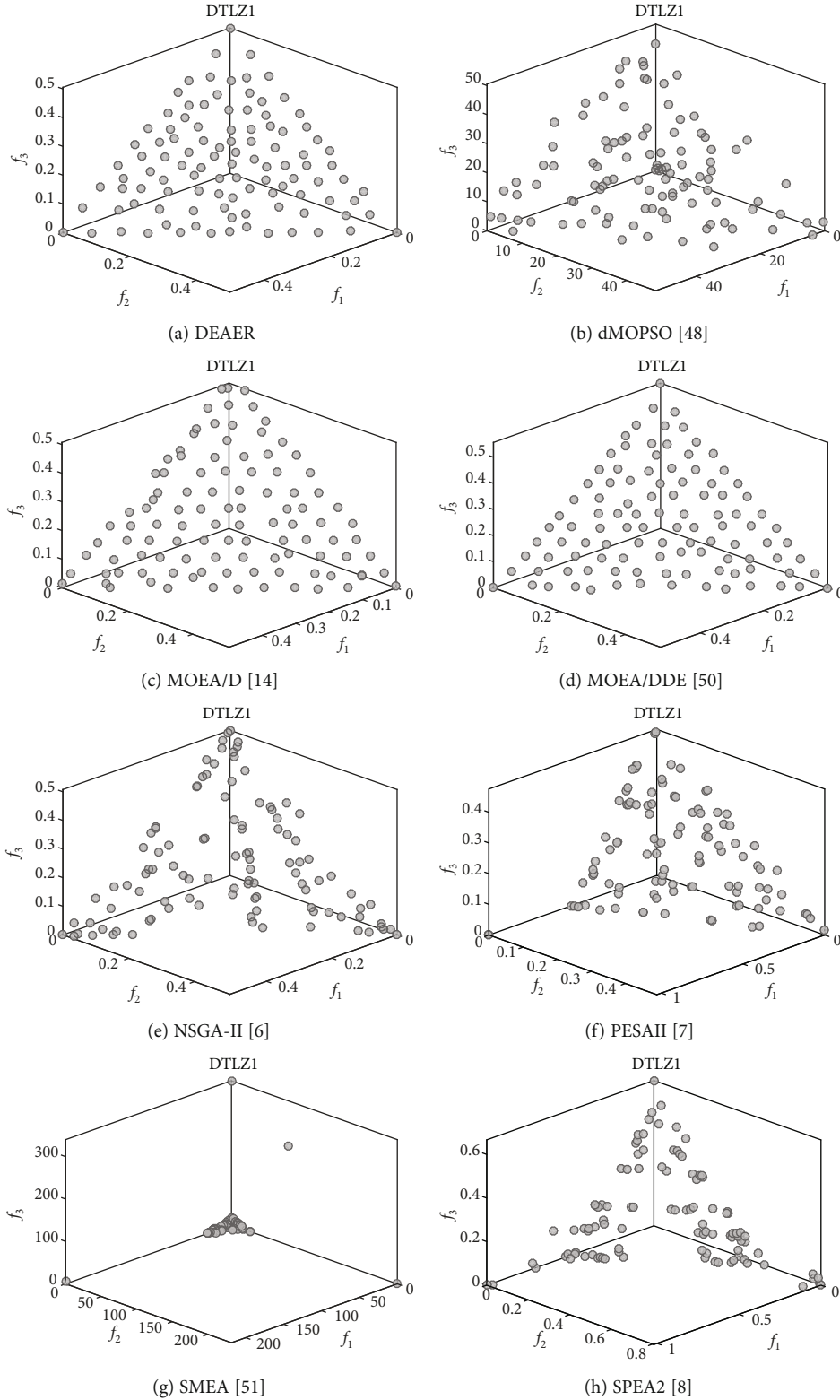


FIGURE 4: The nondominated solutions obtained by DERE, dMOPSO [48], MOEA/D [14], MOEA/DDE [50], NSGA-II [6], PESAI [7], SMEA [51], and SPEA2 [8] on DTLZ1 problem with 3-objective.

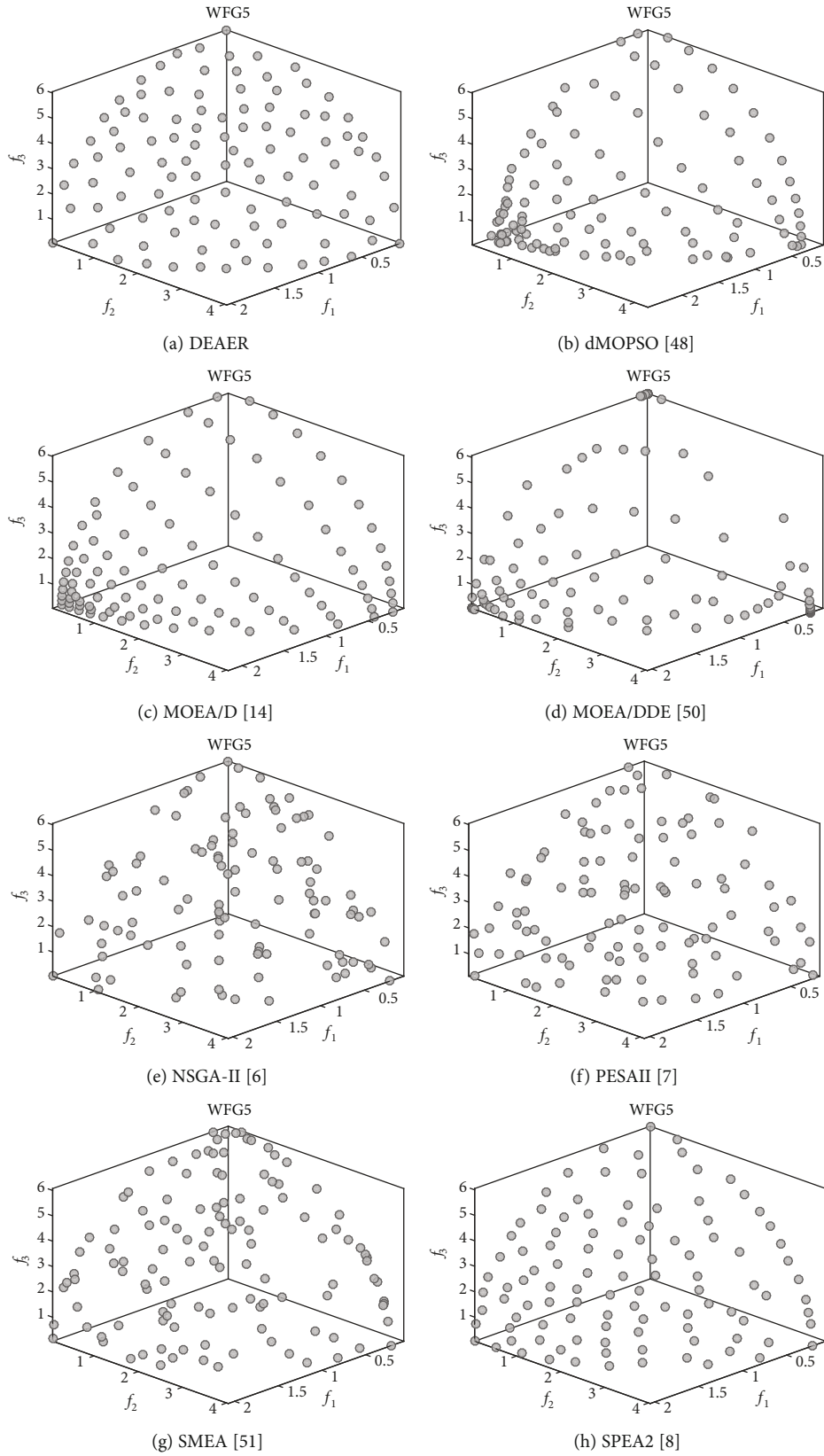


FIGURE 5: The nondominated solutions obtained by DERA, dMOPSO [48], MOEA/D [14], MOEA/DDE [50], NSGA-II [6], PESAI [7], SMEA [51], and SPEA2 [8] on WFG5 problem with 3-objective.

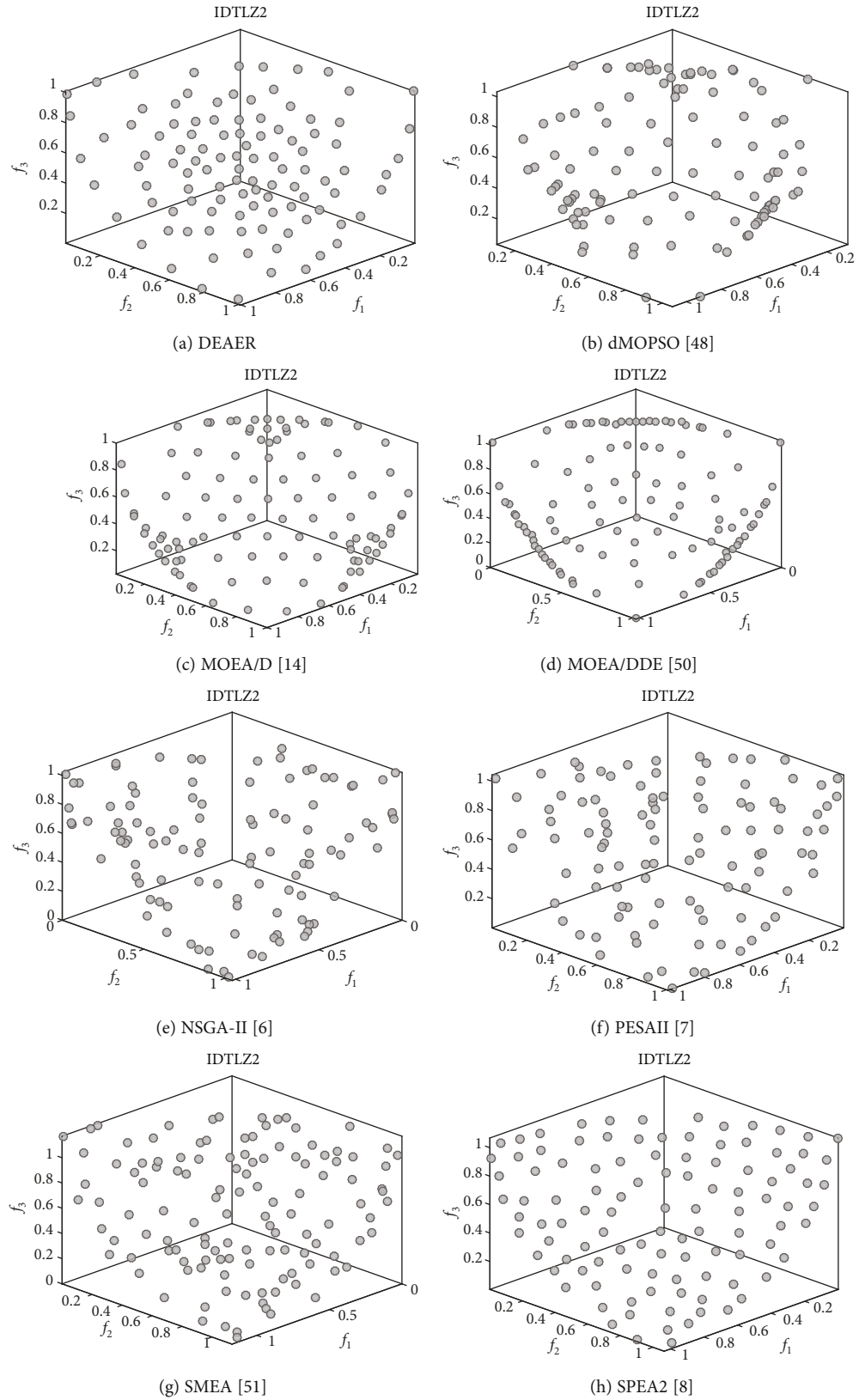


FIGURE 6: The nondominated solutions obtained by DERE, dMOPSO [48], MOEA/D [14], MOEA/DDE [50], NSGA-II [6], PESAI [7], SMEA [51], and SPEA2 [8] on IDTLZ2 problem with 3-objective.



TABLE 4: The statistical results (mean and standard deviation) of the IGD values obtained by DERE, PESAI, dMOPSO, NSGA-II, and MOEA/DDE on UF1-UF10. The best results are highlighted in blue.

Problem	$M$	DEREA	PESAI [7]	dMOPSO [48]	NSGA-II [6]	MOEA/DDE [50]
UF1	2	7.6648e-2 (1.86e-2)	1.2949e-1 (4.29e-2)-	2.1650e-1 (6.57e-2)-	8.2028e-2 (1.29e-2) $\approx$	3.8810e-2 (1.64e-2)+
UF2	2	2.4740e-2 (4.42e-3)	3.7279e-2 (1.57e-2)-	5.3844e-2 (6.62e-3)-	2.7868e-2 (6.93e-3) $\approx$	2.1435e-2 (8.96e-3)+
UF3	2	2.1154e-1 (4.33e-2)	2.5232e-1 (4.31e-2)-	3.3960e-1 (2.08e-2)-	2.3749e-1 (4.41e-2)-	2.1562e-1 (4.37e-2) $\approx$
UF4	2	4.0734e-2 (1.72e-3)	4.6986e-2 (1.70e-3)-	9.7788e-2 (7.03e-3)-	4.2393e-2 (1.25e-3)-	5.9141e-2 (5.89e-3)-
UF5	2	2.7954e-1 (6.80e-2)	3.9438e-1 (1.06e-1)-	1.1205e+0 (2.25e-1)-	3.2780e-1 (8.90e-2)-	4.5750e-1 (1.02e-1)-
UF6	2	2.0051e-1 (7.66e-2)	3.1073e-1 (1.24e-1)-	9.4042e-1 (7.82e-2)-	2.0025e-1 (8.76e-2) $\approx$	3.1438e-1 (1.52e-1)-
UF7	2	8.6449e-2 (1.17e-1)	2.5804e-1 (1.88e-1)-	1.1975e-1 (2.68e-2)-	1.0507e-1 (1.34e-1) $\approx$	2.2722e-2 (4.22e-2)+
UF8	3	1.4673e-1 (4.44e-2)	2.2468e-1 (5.76e-2)-	2.9173e-1 (3.77e-2)-	2.4555e-1 (5.27e-2)-	1.3471e-1 (3.73e-2) $\approx$
UF9	3	1.3775e-1 (5.50e-2)	3.3876e-1 (1.27e-1)-	4.1963e-1 (5.45e-2)-	3.9328e-1 (1.40e-1)-	1.9027e-1 (5.71e-2)-
UF10	3	3.3628e-1 (4.64e-2)	4.4792e-1 (6.74e-2)-	8.5001e-1 (2.25e-1)-	4.0483e-1 (7.23e-2)-	5.4458e-1 (5.32e-2)-
		+/- / $\approx$	0/10/0	0/10/0	0/10/0	3/5/2

' + ', ' - ' and '  $\approx$  ' indicates that the result is significantly better, significantly worse, and statistically similar to that obtained by DERE, respectively.

MOEAs estimate the density of each individual in the population based on the position of the individual's neighbors. Therefore, the long-term existence of low-convergence individuals will cause the disappearance of some high-convergence individuals in the limitation of population size in the environmental selection. When estimating the surrounding area density around in any individual in a population, reflecting the relative proximity of the individual to the PF by DER will shift the position of other individuals based their convergence. This will better maintain the diversity of solutions. Hence, DERE adopts DER strategy in order to select the convergence of the individual in the population.

### 3. Experimental Results and Analysis

In this section, the suggested DERE with some state-of-the-art MOEAs was designed for solving MOPs to demonstrate the performance of the DER, and these classical MOEAs include GDE3 [47], dMOPSO [48], MOPSO [49], MOEA/D [14], MOEA/DDE [50], NSGA-II [6], PESAI [7], SMEA [51], and SPEA2 [8]. The main step of the experimental process includes

- (1) Firstly, the experimental settings used in this paper are illustrated
- (2) Secondly, each algorithm is run 30 times independently for each test question, and the Wilcoxon rank sum test is adopted to compare the results obtained by DERE and those by others compared algorithms at a significance level of 0.05

In the experiments, test problems include multiple widely used test suites that are employed and DTLZ1-DTLZ7, IDTLZ1, IDTLZ2 [30], WFG1-WFG9, ZDT1-ZDT6, and UF1-UF9, where seven test problems are DTLZ1-DTLZ7 from the DTLZ test suite [49], test problems WFG1-WFG9 and UF1-UF9 taken from WFG test suite [22], and UF test suites [52], respectively. Six test problems are ZDT1-ZDT6 that are taken from ZDT [14] test suites. For IDTLZ1,

IDTLZ2, DTLZ1-DTLZ7, and MaF1-MaF7, the number of decision variables is set to  $n = k + M - 1$ , where  $M$  is the objective number,  $k = 5$  is used for IDTLZ1 and DTLZ1,  $k = 10$  is used for IDTLZ2, and DTLZ2-DTLZ6,  $k = 20$ , is used for DTLZ7. For WFG1-WFG9, the number of decision variables is set to  $n = k + l$ ,  $k$  is used for  $(M - 1)$ , and  $l$  is used for 10. Pareto fronts of the above part test instance are given in Table 1. For UF test suites and ZDT test suites, UF1-UF10 and ZDT1-ZDT3 test instance in the number of decision variables is set to 30, and ZDT4 with ZDT6 and ZDT5 corresponds to the setting  $n = 10$  and  $n = 80$ , where UF test suites contain 10 continuous multiobjective optimization problems with constraints. All the compared MOEAs in this paper are implemented based on MOEA platform PlatEMO [53].

#### 3.1. Experimental Setting

- (1) Setting for crossover and mutation operators: The simulated polynomial mutation [40] and binary crossover (SBX) [41] are applied in all MOEAs. The probabilities of crossover and mutation are set to 1.00 and  $1/D$  (where  $D$  denoting the number of decision variables). The distribution indexes of both SBX and polynomial mutation are set to 20
- (2) Population size: The population sizes of the algorithm for experimental comparison are determined by the number of reference vectors or reference points in this paper. For problems with  $M \geq 3$ , the population size is generated by a two-layer vector generation strategy [54]. For the accuracy of the comparison results, the specific setting of the population size of the comparison algorithms is summarized in Table 2, and the population sizes are also set according to Table 2 for the proposed DERE
- (3) Performance metrics: In these experiments, the widely used hypervolume (HV) and the inverted generational distance (IGD) were adopted to as

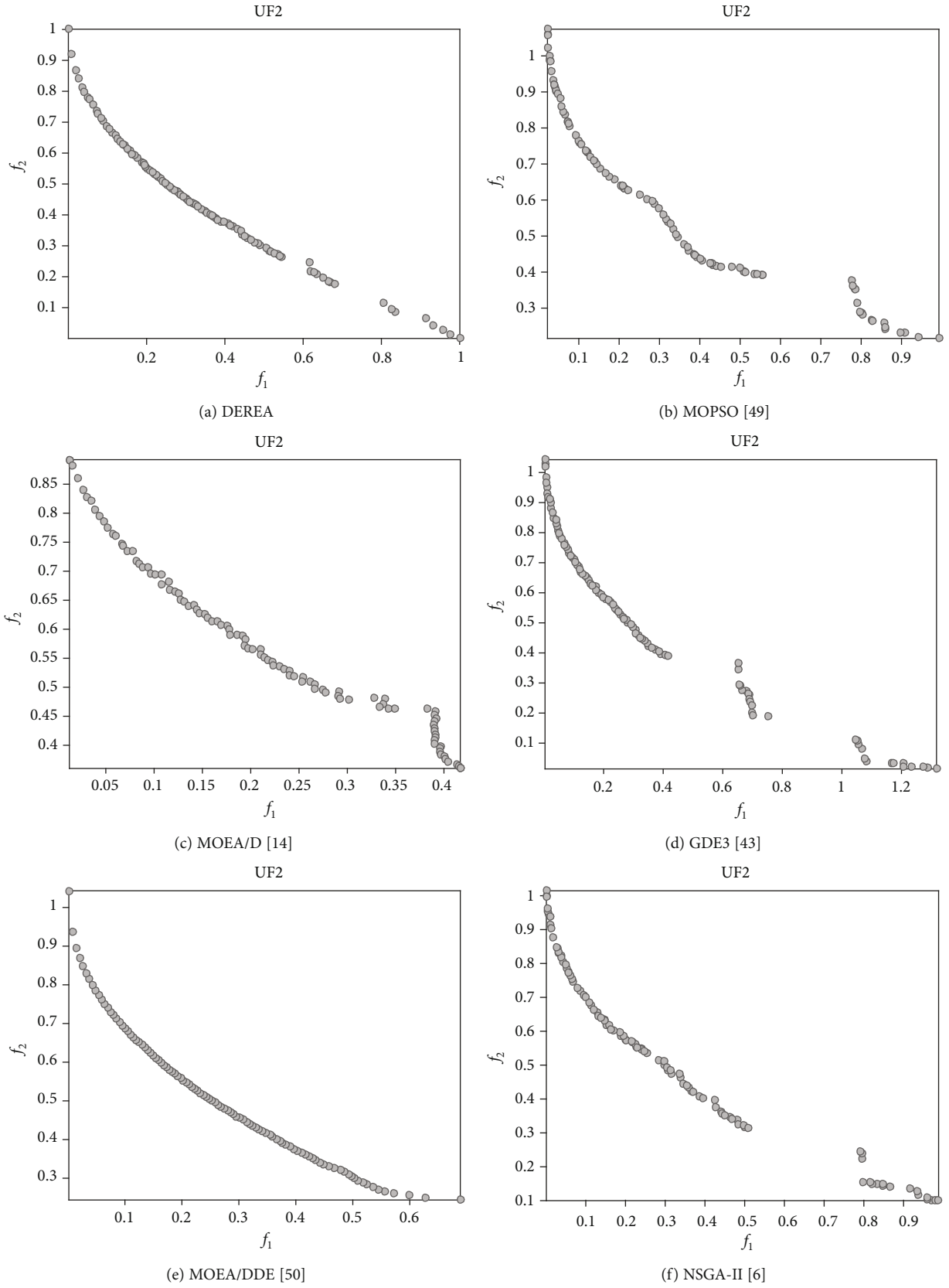


FIGURE 7: Continued.

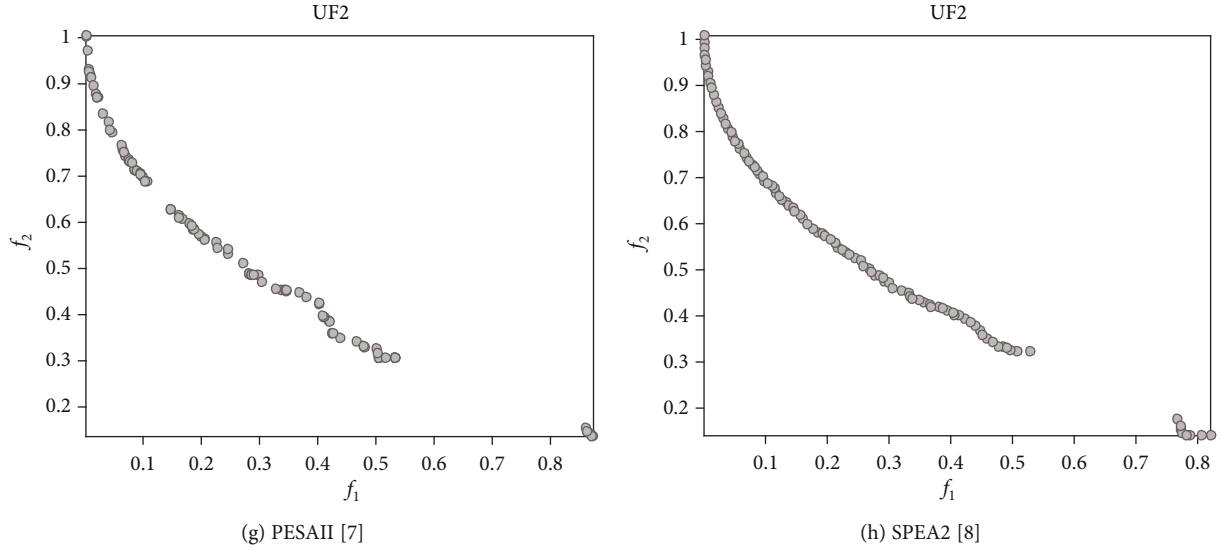


FIGURE 7: The nondominated solutions obtained by DERE, MOPSO, MOEA/D, GDE3, MOEA/DDE, NSGA-II, PESAI, and SPEA2 on UF2 problem with 2-objective.

TABLE 5: The statistical results (mean and standard deviation) of the IGD values obtained by DERE, PESAI, dMOPSO, NSGA-II, and MOEA/D on ZDT1-ZDT6. The best results are highlighted in blue.

Problem	$M$	DERE	PESAI [7]	dMOPSO [48]	NSGA-II [6]	MOEA/D [14]
ZDT1	2	4.1510e-3 (7.97e-5)	1.1847e-2 (5.24e-3)-	8.5475e-3 (2.98e-3)-	4.6078e-3 (1.47e-4)-	4.6049e-3 (6.45e-4)-
ZDT2	2	4.6195e-3 (1.98e-5)	1.0719e-2 (1.44e-3)-	8.5282e-2 (2.06e-1)-	5.0609e-3 (2.40e-4)-	3.8946e-3 (6.24e-5)+
ZDT3	2	2.0741e-2 (3.25e-2)	6.5135e-2 (6.04e-2)-	3.1740e-2 (9.28e-2)-	3.0418e-2 (3.52e-2)-	2.4470e-2 (2.07e-2)-
ZDT4	2	4.8255e-3 (6.43e-4)	1.1187e-2 (1.86e-3)-	7.3167e-3 (2.21e-3)-	4.8779e-3 (5.48e-4)≈	1.0252e-2 (2.48e-3)-
ZDT5	2	1.1021e+0 (2.90e-1)	4.9351e-1 (2.36e-1)≈	0.0000e+0 (0.00e-0)-	1.7390e-2 (3.69e-2)≈	1.2512e+1 (3.96e-1)-
ZDT6	2	3.3452e-3 (5.76e-5)	8.0858e-3 (1.29e-3)-	3.1043e-3 (2.03e-6)+	3.7711e-3 (1.43e-4)-	5.5360e-3 (8.50e-4)-
		+/- /≈	0/5/1	1/5/0	0/4/2	1/5/0

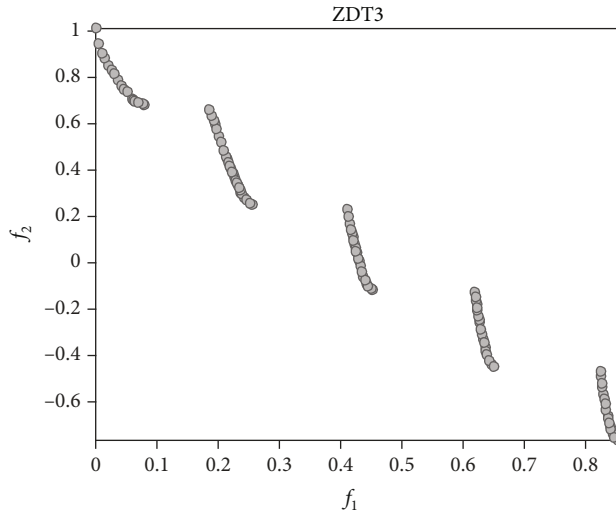
'+' , '-' and '≈' indicates that the result is significantly better, significantly worse, and statistically similar to that obtained by DERE, respectively.

performance metrics for comparisons among the experiments results obtain by each algorithm. All the objectives are normalized according to the ideal point and the worst point of the Pareto optimal front before HV calculation, then calculated with a reference point (1.1, 1.1, ..., 1.1). Besides, in the calculation of IGD metric, almost 5,000 uniformly distributed points are sampled on the PF by Das and Dennis's method for each test instance, where defined by the symbols '+', '-', and '≈' indicate that the performance metrics by others MOEA is significantly better, significantly worse, and statistically similar to that obtained by DERE, respectively

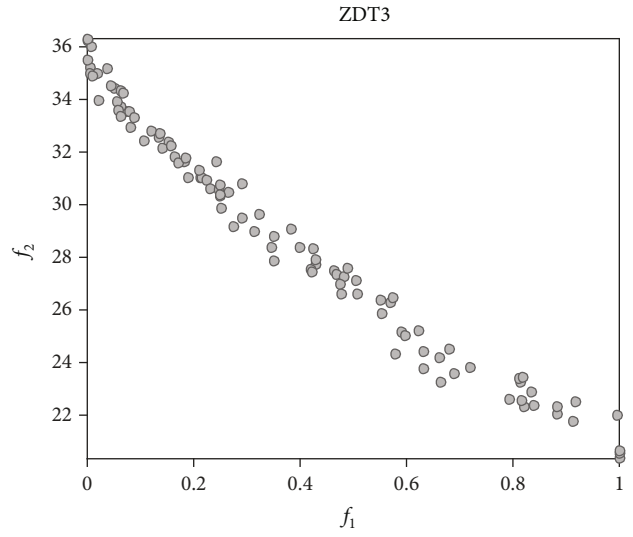
- (4) Termination condition: The termination criterion of each run is the maximal number of generations. For all test problems adopted in the experiment, set the maximal number of generations based on test problems or different number objective. For DTLZ1, DTLZ3, and WFG1-WFG9, the maximal number of generations is set to 1000. For DTLZ2, DTLZ4, and UF test suite, the maximal number of genera-

tions is set to 500. For ZDT1-ZDT6 test suite, the maximal number of generations is set 300. The maximal number of generations of each algorithm to deal with identical problems is the same

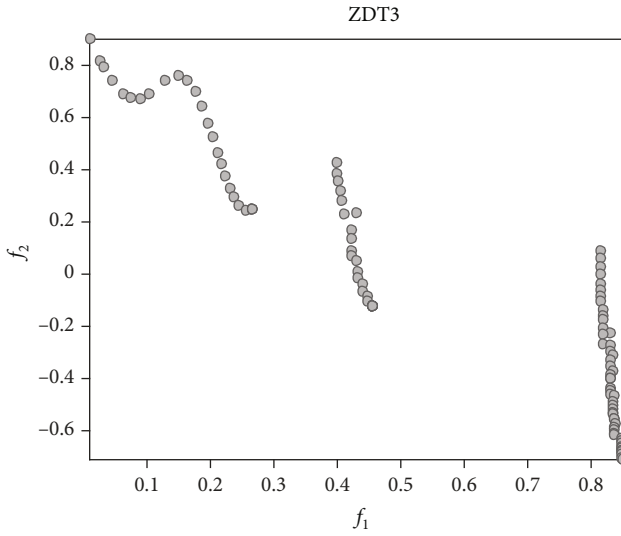
- (5) Specific parameter settings in the compared algorithm: For PESA-II, the number of divisions in each objective is set to 10. For MOEA/D and MOEA/DDE, the size of weight vectors neighborhood  $T$  is set to  $[0.1 \times N]$  (where  $N$  is the population size), the maximum number of solutions substitute by each offspring  $n_r$  is set to  $[0.01 \times N]$ , and the Tchebycheff method is employed as the aggregation function. In addition, for MOEA/DDE, the probability of selecting neighborhood solutions  $\delta$  is set to 0.9. For dMOPSO,  $T_\alpha$  denotes the age threshold that is set to 2. For SMEA, the number of neurons in each dimension of the latent space  $D$  is determined by the number of dimensions  $M$ , the initial self-organizing maps learn rate  $\tau_0$ , and the size of neighborhood mate pools  $H$  are set to 0.7 and 5, respectively. In the experimental comparisons, for GDE3,



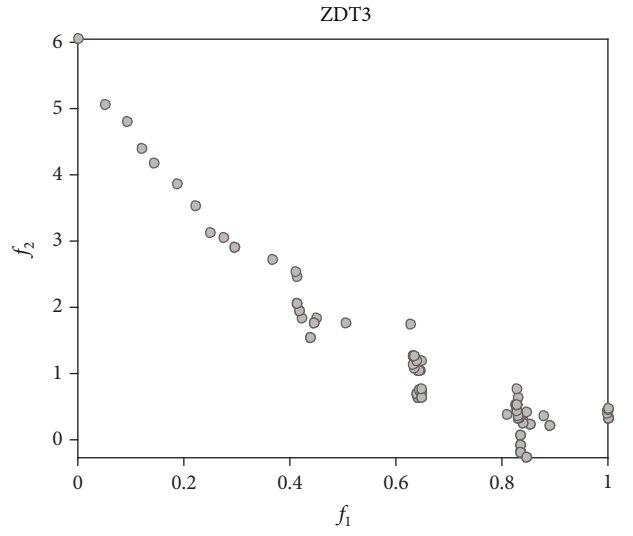
(a) DERE



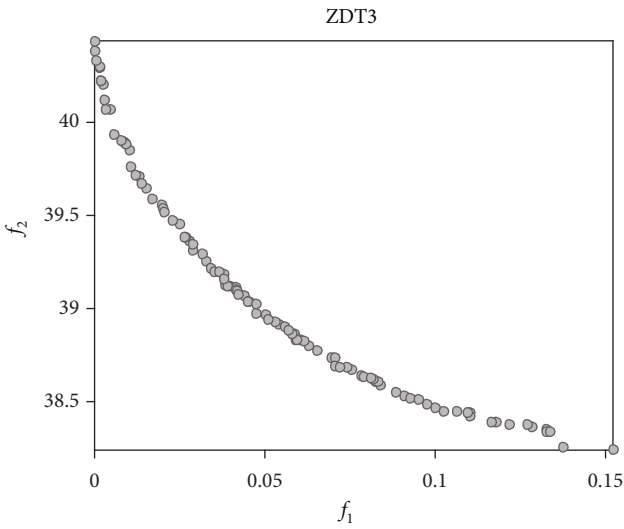
(b) MOPSO [49]



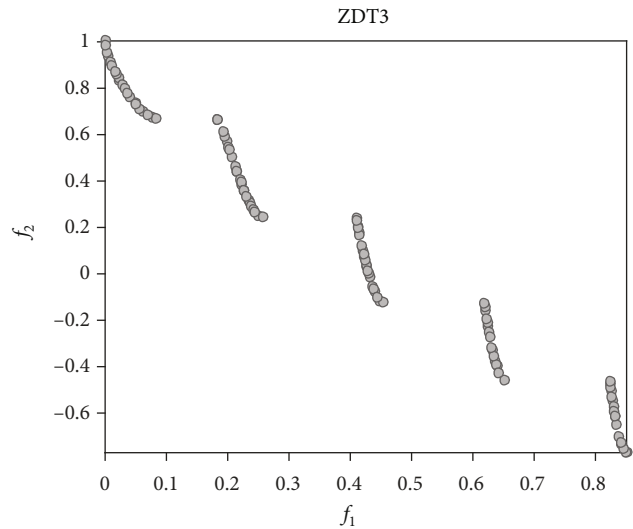
(c) MOEA/D [14]



(d) GDE3 [43]



(e) MOEA/DDE [50]



(f) NSGA-II [6]

FIGURE 8: Continued.

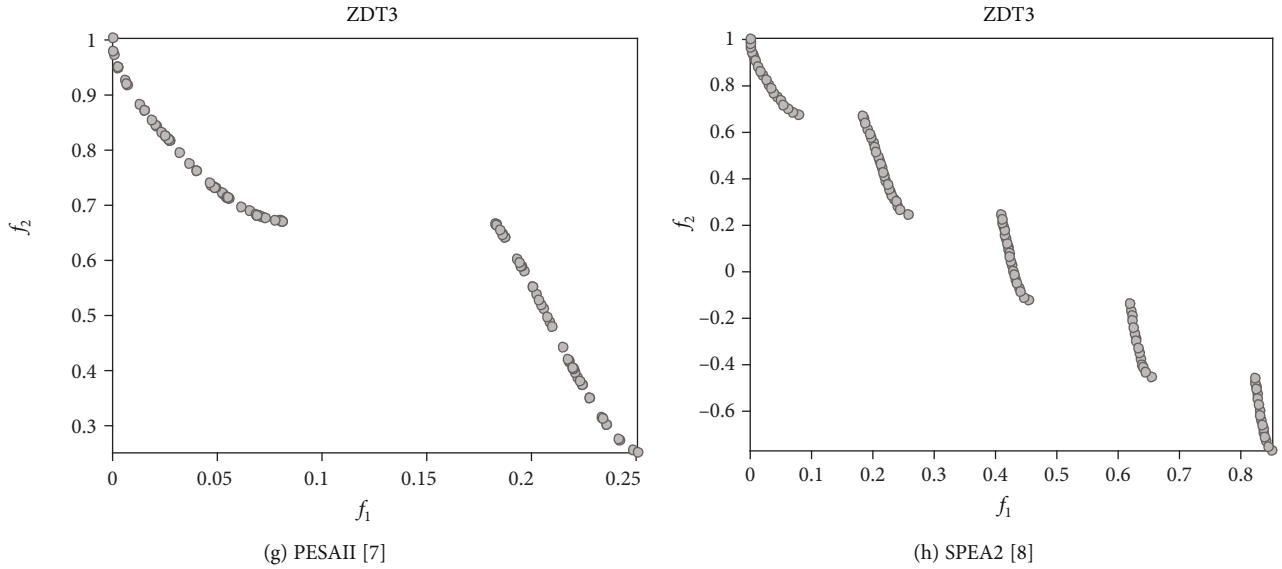


FIGURE 8: The nondominated solutions obtained by DERE, MOPSO, MOEA/D, GDE3, MOEA/DDE, NSGA-II, PESAI, and SPEA2 on ZDT3 problem with 2-objective.

NSGA-II, DERE, and SPEA2, there are no additional specific additional parameter

### 3.2. Comparisons between DERE and Existing MOEAs for Solving MOPs

**3.2.1. Performance on WFG1–WFG9, DTLZ1–DTLZ7, and IDTLZ1–IDTLZ2.** Table 3 demonstrates the HV values obtained by DERE and other four classical MOEAs designed for solving MOPs, namely, dMOPSO, MOEA/D, NSGA-II, and SPEA2 on WFG1 to WFG9 and DTLZ1 to DTLZ7, IDTLZ1, and IDTLZ2 with two and three objectives. In general, the suggested DERE significantly outperforms the other four MOEAs in term of the HV values. For test problems with irregular PF, Table 3 summarizes the HV values of each algorithm solving the test problems, and DERE shows a performance surpass others competitors on the majority of the test instances. However, the performance of DERE shows significance advantages with dMOPSO, MOEA/D, NSGA-II, and SPEA2 on the regular PFs. NSGA-II is outperformance by DERE on more than 24 test instances, but it performs pretty well on WFG1 and IDTLZ1. MOEA/D and SPEA2 achieve competitive results on IDTLZ problems. The dMOPSO does not have an advantage and is surpassed by DERE on almost all the test problems. The performance will become poor when algorithms solving with irregular PF test problems, and algorithms with density estimation (i.e., DERE) indicate clear improvement over the others competitor MOEAs. However, the performance of MOEA/D, NSGA-II, and SPEA2 deteriorates significantly, being inferior to both DERE on the most test instance. This phenomenon resulted from the fact that the Pareto fronts of these test instances are complicated, disconnected, or degenerate.

To visually understand the experimental results, Figure 4 plots the solution sets on the 3-objective DTLZ1. SMEA per-

forms poorly on this test problem, and the selection solutions do not converge to PF. The other several MOEAs indicate high performance in terms of converges. MOEA/D and MOEA/DDE reach the PF sparse area and obtain a distributed solution sets, but its selection solutions may not be close to the true PF, which illustrates why it has small HV value. The solutions are uniformly distributed on PF by DERE obtained. It is easy to see that the solution obtained by dMOPSO, NSGA-II, PESAI, and SPEA2 has poor performance, which due to solutions located near the PF boundaries. Figure 5 plots the final solutions on the 3-objective WFG5. NSGA-II, PESAI, and SPEA2 converge to the PF region but its solutions fail to uniformly distribute. Due to some dominant solutions that were deleted with that algorithm selective solution only by individual position information, SMEA indicates poor performance in terms of coverage. The other three MOEAs have their own traits in the distribution of the solutions, and they obtain smaller HV values. dMOPSO selected optimal solutions of away from the PF and MOEA/D obtain solutions located near the PF boundaries. And MOEA/DDE has poor performance with diversity.

The final solutions on the 3-objective IDTLZ2 problem plotted in three-dimensional space are provided in Figure 6. dMOPSO, NSGA-II, PESAI, and SMEA all converge to the PF area, and however, fail to manager population diversity. The other three competitors indicate clear advantages in preserving diversity. These algorithms show performance that are similar to DERE, but the best HV value by DERE was obtained.

**3.2.2. Performance on UF1–UF10.** The results on problems with UF in terms of IGD are presented in Table 4. DERE performs the best among the five algorithms. It obtains the best results on ten test instances. NSGA-II and MOEA/DDE are also competitive with DERE on the majority of the test instances, i.e., UF1, UF2, and UF7. Other competitors (i.e.,

PESAI and dMOPSO) perform inferior to the above algorithms, where PESAI and dMOPSO indicate performance poorly on the most test instances.

To further observe the differences among solutions set obtained by the several compared algorithms, add three classical algorithms (i.e., MOEA/D, GDE3 and SPEA2). Figure 7 depicts the objective values of nondominated solution sets with the IGD value. It can be clearly seen that the solution sets obtained by DERE and MOEA/DDE have shown good performance, and this is why IGD values are similar. The performance of other algorithms is similar, but except MOEA/D, which it selected optimal solutions that fail converge to the PF. Comparing this algorithm in UF2 instance DERE has indicated a slight advantage.

**3.2.3. Performance on ZDT1–ZDT6.** Table 5 presents the mean and standard deviation of the IGD values of the final solutions by each algorithm obtained for ZDT test suites. Table 5 reveals that in terms of IGD, the final solutions by DERE are better than the other algorithms for ZDT1, ZDT3, ZDT4, and ZDT5 instances. DERE performance surpasses the other four algorithms.

To facilitate visual comparison, Figure 8 shows the final solutions of a run of the several algorithms. Clearly, the solutions of DERE have good balance between diversity and convergence. MOEA/D and PESAI struggle to maintain diversity, but lead to its selected solutions concentrated in a local optimum. NSGA-II and SPEA2 selected solutions show good diversity, but convergence performs slightly worse DERE. However, MOPSO and GDE3 indicate poor performance, and the most of the solutions have a poor convergence, thus leading to its high IGD value.

## 4. Conclusion

This paper proposes density estimation ranking strategy for multiobjective optimization, namely, DERE. In this paper, a solution ranking method named DER is proposed. The DER is introduced to adapt different shapes of Pareto fronts. In addition, the DER is the most significant contribution of this paper, changing the position of individuals by their relative proximity to the Pareto front, DER considers the convergence and diversity of each individual in the population, and the implementation of it is very simple.

Through extensive comparison of several widely use test suites, a systematic experiment was carried out. From the experiment results, the DERE is quite competitive with the several state-of-the-art algorithms on a majority of the test suites. Moreover, due to DER are employed, the performance of DERE is less dependent on the Pareto front shapes and is robust in solving problems having irregular Pareto fronts.

In the future, we will have a future investigation into the density estimation mechanism in the selected strategy. We also want the future how to deal with more challenging problems, some of which are not easy to obtain during the search process. And it is also meaningful to study many objective optimization problems.

## Data Availability

The dataset used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors have declared that no competing interests exist.

## References

- [1] E. J. S. Pires, P. B. de Moura Oliveira, and J. A. T. Machado, "Multi-objective Genetic Manipulator Trajectory Planner," in *Workshops on Applications of Evolutionary Computation*, Springer, Berlin, Heidelberg, 2004.
- [2] C. Wang, Z. Ye, and W. Wang, "A Multi-Objective Optimization and Hybrid Heuristic Approach for Urban Bus Route Network Design," *IEEE Access*, vol. 8, pp. 12154–12167, 2020.
- [3] A. Asrari, S. Lotfifard, and M. S. Payam, "Pareto dominance-based multiobjective optimization method for distribution network reconfiguration," *IEEE Transactions on Smart Grid*, vol. 7, no. 3, pp. 1401–1410, 2016.
- [4] Z. Zhao, F. Lin, X. Wang, and R. Liu, "Electricity price decision-making method of electricity selling company based on multi-objective optimization and min-max regret theory," *Journal of Physics: Conference Series*, vol. 1650, no. 3, article 032171, 2020.
- [5] D. W. Corne and J. D. Knowles, "Techniques for highly multi-objective optimisation: some nondominated points are better than others," in *Proceedings of the International Conference on Genetic and Evolutionary Computation*, pp. 773–780, London, England, 2007.
- [6] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002.
- [7] D. Corne, N. Jerram, J. D. Knowles, M. Oates, and J. Martin, "PESA-II: Region-based Selection in Evolutionary Multiobjective Optimization," in *Proceedings of the 3rd annual conference on genetic and evolutionary computation*, pp. 283–290, San Francisco, CA, USA, 2001.
- [8] E. Zitzler, M. Laumanns, and L. Thiele, "SPEA2: Improving the strength Pareto evolutionary algorithm," Tech. Rep. TIK-Report 103, ETH Zurich, 2001.
- [9] K. Deb, M. Mohan, and S. Mishra, "Towards a Quick Computation of Well-Spread Pareto-Optimal Solutions," in *Evolutionary Multi-Criterion Optimization, Second International Conference, Emo, Faro, Portugal, April*. Springer-Verlag, 2003.
- [10] C. Yue, B. Qu, and L. Jing, "A Multiobjective particle swarm optimizer using ring topology for solving multimodal Multi-objective problems," *IEEE Transactions on Evolutionary Computation*, vol. 22, no. 5, pp. 805–817, 2018.
- [11] T. Cheng, M. Chen, P. J. Fleming, Z. Yang, and S. Gan, "A novel hybrid teaching learning based multi-objective particle swarm optimization," *Neurocomputing*, vol. 222, pp. 11–25, 2016.
- [12] H. L. Liu, F. Gu, and Q. Zhang, "Decomposition of a multiobjective optimization problem into a number of simple multi-objective sub-problems," *IEEE Transactions on Evolutionary Computation*, vol. 18, no. 3, pp. 450–455, 2014.
- [13] T. Murata, "Specification of genetic search directions in cellular multi-objective genetic algorithms," in *International*

- Conference on Evolutionary Multi-Criterion Optimization*, pp. 82–95, Berlin, Heidelberg, 2001.
- [14] Q. Zhang and H. Li, “MOEA/D: a multiobjective evolutionary algorithm based on decomposition,” *IEEE Transactions on Evolutionary Computation*, vol. 11, no. 6, pp. 712–731, 2008.
- [15] R. Cheng, Y. Jin, M. Olhofer, and B. Sendhoff, “A reference vector guided evolutionary algorithm for many-objective optimization,” *IEEE Transactions on Evolutionary Computation*, vol. 20, no. 5, pp. 773–791, 2016.
- [16] X. He, Y. Zhou, Z. Chen, and Q. Zhang, “Evolutionary many-objective optimization based on dynamical decomposition,” *IEEE Transactions on Evolutionary Computation*, vol. 23, no. 3, pp. 361–375, 2019.
- [17] K. Deb and H. Jain, “An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part I: solving problems with box constraints,” *IEEE Transactions on Evolutionary Computation*, vol. 18, no. 4, pp. 577–601, 2014.
- [18] R. Cheng, Y. Jin, K. Narukawa, and B. Sendhoff, “A multiobjective evolutionary algorithm using Gaussian process-based inverse modeling,” *IEEE Transactions on Evolutionary Computation*, vol. 19, no. 6, pp. 838–856, 2015.
- [19] G. Li, W. Wang, W. Zhang, Z. Wang, H. Tu, and W. You, “Grid search based multi-population particle swarm optimization algorithm for multimodal multi-objective optimization,” *Swarm and Evolutionary Computation*, vol. 62, article id 100843, 2021.
- [20] M. Wu, K. Li, S. Kwong, Q. Zhang, and J. Zhang, “Learning to decompose: a paradigm for decomposition-based multiobjective optimization,” *IEEE Transactions on Evolutionary Computation*, vol. 23, no. 3, pp. 376–390, 2019.
- [21] H. L. Liu, F. Gu, and Q. Zhang, “Decomposition of a multiobjective optimization problem into a number of simple multiobjective subproblems,” *IEEE Transactions on Evolutionary Computation*, vol. 18, no. 3, pp. 450–455, 2014.
- [22] S. Jiang and S. Yang, “A strength Pareto evolutionary algorithm based on reference direction for multi-objective and many-objective optimization,” *IEEE Transactions on Evolutionary Computation*, vol. 21, no. 3, pp. 329–346, 2017.
- [23] N. Beume, B. Naujoks, and M. Emmerich, “SMS-EMOA: multiobjective selection based on dominated hypervolume,” *European Journal of Operational Research*, vol. 181, no. 3, pp. 1653–1669, 2007.
- [24] A. Menchaca-Mendez and C. Coello, “GDE-MOEA: A new MOEA based on the generational distance indicator and  $\epsilon$ -dominance,” in *2015 IEEE Congress on Evolutionary Computation (CEC)*, pp. 947–955, Sendai, Japan, 2015.
- [25] A. F. Ionescu, “Multi-objective Evolutionary Algorithms: Decomposition Versus Indicator-Based Approach,” in vol. 404 of *Studies in Fuzziness and Soft Computing*, Springer, Cham, 2021.
- [26] K. Li, S. Kwong, J. Cao, M. Li, J. Zheng, and R. Shen, “Achieving balance between proximity and diversity in multi-objective evolutionary algorithm,” *Information Sciences*, vol. 182, no. 1, pp. 220–242, 2012.
- [27] Y. Sun, G. G. Yen, and Z. Yi, “IGD Indicator-based evolutionary algorithm for many-objective optimization problems,” *IEEE Transactions on Evolutionary Computation*, vol. 23, no. 2, pp. 173–187, 2019.
- [28] R. Hernandez Gomez and C. A. Coello, “Improved metaheuristic based on the R2 indicator for many-objective optimization,” in *Proceedings of the 2015 International Conference on Genetic and Evolutionary Computation*, pp. 679–686, New York, NY, USA, 2015.
- [29] J. Bader and E. Zitzler, “HypE: An algorithm for fast Hypervolume-based many-objective optimization,” *Evolutionary Computation*, vol. 19, no. 1, pp. 45–76, 2011.
- [30] Y. Tian, R. Cheng, X. Zhang, F. Cheng, and Y. Jin, “An Indicator Based Multi-Objective Evolutionary Algorithm with Reference Point Adaptation for Better Versatility,” *IEEE Transactions on Evolutionary Computation*, vol. 22, pp. 609–622, 2017.
- [31] X. Ma, F. Liu, Y. Qi et al., “A multiobjective evolutionary algorithm based on decision variable analyses for multiobjective optimization problems with large-scale variables,” *IEEE Transactions on Evolutionary Computation*, vol. 20, no. 2, pp. 275–298, 2016.
- [32] H. Li, Q. Zhang, and J. Deng, “Biased multiobjective optimization and decomposition algorithm,” *IEEE Transactions on Cybernetics*, vol. 47, no. 1, pp. 52–66, 2017.
- [33] G. Chen and J. Li, “A diversity ranking based evolutionary algorithm for multi-objective and many-objective optimization,” *Swarm and Evolutionary Computation*, vol. 48, pp. 274–287, 2019.
- [34] Y. Liang, H. Huang, Z. Cai, and Z. Hao, “Multiobjective evolutionary optimization based on fuzzy multicriteria evaluation and decomposition for image matting,” *IEEE Transactions on Fuzzy Systems*, vol. 27, no. 5, pp. 1100–1111, 2019.
- [35] H. Wang, L. Jiao, and X. Yao, “Two\_Arch2: An improved Two-archive algorithm for many-objective optimization,” *IEEE Transactions on Evolutionary Computation*, vol. 19, no. 4, pp. 524–541, 2015.
- [36] Y. Liu, D. Gong, J. Sun, and Y. Jin, “A many-objective evolutionary algorithm using a one-by-one selection strategy,” *IEEE Transactions on Cybernetics*, vol. 47, no. 9, pp. 2689–2702, 2017.
- [37] M. Li, S. Yang, and X. Liu, “Pareto or non-Pareto: bi-criterion evolution in multiobjective optimization,” *IEEE Transactions on Evolutionary Computation*, vol. 20, no. 5, pp. 645–665, 2016.
- [38] H. Ishibuchi, Y. Setoguchi, H. Masuda, and Y. Nojima, “Performance of decomposition-based many-objective algorithms strongly depends on Pareto front shapes,” *IEEE Transactions on Evolutionary Computation*, vol. 21, no. 2, pp. 169–190, 2017.
- [39] C. M. Anderson-Cook, “Practical Genetic Algorithms,” *Journal of the American Statistical Association*, vol. 100, no. 471, pp. 1099–1099, 2005.
- [40] K. Deb and M. Goyal, “A combined genetic adaptive search (gene AS) for engineering design,” *Computer Science and Informatics*, vol. 26, no. 4, pp. 30–45, 1996.
- [41] K. Deb and R. B. Agrawal, “Simulated binary crossover for continuous search space,” *Complex Systems*, vol. 9, no. 4, pp. 115–148, 1995.
- [42] C. A. Coello Coello and E. Mezura Montes, “Constraint-handling in genetic algorithms through the use of dominance-based tournament selection,” *Advanced Engineering Informatics*, vol. 16, no. 3, pp. 193–203, 2002.
- [43] G. G. Yen and Haiming Lu, “Dynamic multiobjective evolutionary algorithm: adaptive cell-based rank and density estimation,” *IEEE Transactions on Evolutionary Computation*, vol. 7, no. 3, pp. 253–274, 2003.

- [44] M. Li, J. Zheng, K. Li, Q. Yuan, and R. Shen, "Enhancing diversity for average ranking method in evolutionary many-objective optimization," in *Parallel Problem Solving from Nature, PPSN XI*, pp. 647–656, Springer, 2010.
- [45] M. Li, J. Zheng, R. Shen, K. Li, and Q. Yuan, "A grid-based fitness strategy for evolutionary many-objective optimization," in *Proceedings of the 12th annual conference on Genetic and evolutionary computation - GECCO '10*, pp. 463–470, New York, NY, USA, 2010.
- [46] M. Li, S. Yang, and X. Liu, "Shift-based density estimation for Pareto-based algorithms in many-objective optimization," *IEEE Transactions on Evolutionary Computation*, vol. 18, no. 3, pp. 348–365, 2014.
- [47] S. Kukkonen and J. Lampinen, "GDE3: The third Evolution Step of Generalized Differential Evolution," in *2005 IEEE Congress on Evolutionary Computation*, vol. 1, pp. 443–450, Edinburgh, UK, September 2005.
- [48] S. Z. Martinez and C. A. C. Coello, "A multi-objective particle swarm optimizer based on decomposition," in *Proceedings of the 13th Annual Conference on Genetic and Evolutionary Computation*, pp. 69–76, New York, NY, USA, 2011.
- [49] C. A. Coello and M. S. Lechuga, "MOPSO: a proposal for multiple objective particle swarm optimization," in *Proceedings of the 2002 Congress on Evolutionary Computation. CEC'02 (Cat. No.02TH8600)*, pp. 1051–1056, Honolulu, HI, USA, 2002.
- [50] H. Li and Q. Zhang, "Multiobjective optimization problems with complicated Pareto sets, MOEA/D and NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 13, no. 2, pp. 284–302, 2009.
- [51] H. Zhang, A. Zhou, S. Song, Q. Zhang, X. Z. Gao, and J. Zhang, "A self-organizing multiobjective evolutionary algorithm," *IEEE Transactions on Evolutionary Computation*, vol. 20, no. 5, pp. 792–806, 2016.
- [52] Q. F. Zhang, W. D. Liu, and H. Li, "The performance of a new version of MOEA/D on CEC09 unconstrained MOP test instances," in *IEEE Congress on Evolutionary Computation*, Trondheim, Norway, 2009.
- [53] Y. Tian, R. Cheng, X. Zhang, and Y. Jin, "PlatEMO: a MATLAB platform for evolutionary multi-objective optimization [educational forum]," *IEEE Computational Intelligence Magazine*, vol. 12, no. 4, pp. 73–87, 2017.
- [54] K. Li, Q. Zhang, and S. Kwong, "An evolutionary many-objective optimization algorithm based on dominance and decomposition," *IEEE Transactions on Evolutionary Computation*, vol. 19, no. 5, pp. 694–716, 2015.



## Research Article

# SOSPCNN: Structurally Optimized Stochastic Pooling Convolutional Neural Network for Tetralogy of Fallot Recognition

Shui-Hua Wang,<sup>1</sup> Kaihong Wu,<sup>2</sup> Tianshu Chu,<sup>3</sup> Steven L. Fernandes,<sup>4</sup> Qinghua Zhou,<sup>1</sup> Yu-Dong Zhang<sup>1,3</sup> , and Jian Sun<sup>2</sup> 

<sup>1</sup>School of Informatics, University of Leicester, Leicester LE1 7RH, UK

<sup>2</sup>The Affiliated Children's Hospital of Nanjing Medical University, Nanjing, China

<sup>3</sup>Nanjing Yirongda Institute of Intelligent Medicine and Additive Manufacturing, Nanjing, China

<sup>4</sup>Department of Computer Science, Design & Journalism, Creighton University, Omaha, Nebraska, USA

Correspondence should be addressed to Yu-Dong Zhang; yudong.zhang@le.ac.uk and Jian Sun; sunjian67@njmu.edu.cn

Received 6 May 2021; Revised 27 May 2021; Accepted 5 June 2021; Published 2 July 2021

Academic Editor: Shan Zhong

Copyright © 2021 Shui-Hua Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Aim.** This study proposes a new artificial intelligence model based on cardiovascular computed tomography for more efficient and precise recognition of Tetralogy of Fallot (TOF). **Methods.** Our model is a structurally optimized stochastic pooling convolutional neural network (SOSPCNN), which combines stochastic pooling, structural optimization, and convolutional neural network. In addition, multiple-way data augmentation is used to overcome overfitting. Grad-CAM is employed to provide explainability to the proposed SOSPCNN model. Meanwhile, both desktop and web apps are developed based on this SOSPCNN model. **Results.** The results on ten runs of 10-fold crossvalidation show that our SOSPCNN model yields a sensitivity of  $92.25 \pm 2.19$ , a specificity of  $92.75 \pm 2.49$ , a precision of  $92.79 \pm 2.29$ , an accuracy of  $92.50 \pm 1.18$ , an F1 score of  $92.48 \pm 1.17$ , an MCC of  $85.06 \pm 2.38$ , an FMI of  $92.50 \pm 1.17$ , and an AUC of 0.9587. **Conclusion.** The SOSPCNN method performed better than three state-of-the-art TOF recognition approaches.

## 1. Introduction

Tetralogy of Fallot (TOF) is a congenital defect that influences normal blood flow through the heart [1]. It is made up of 4 defects of the heart and its blood vessels [2]: (a) ventricular septal defect, (b) overriding aorta, (c) right ventricular outflow tract stenosis, and (d) right ventricular hypertrophy. Defects of TOF can cause oxygen in the blood that flows to the rest of the body to be reduced. Infants with TOF have a bluish-looking skin color [3] since their blood does not carry enough oxygen.

Traditional diagnosis of TOF is after a baby is born, often after the infant had an episode of cyanosis during crying or feeding. The most common test is an echocardiogram [4], an ultrasound of the heart that can show problems with the heart structure and how well the heart is working with this defect. Recently, computed tomography (CT) has shown its success in the differential diagnosis of TOF [5], since it can

provide detailed images of many types of cardiovascular issue; besides, computed tomography (CT) can be performed even if the subject has an implanted medical device, unlike magnetic resonance imaging (MRI) [6].

Manual diagnosis on CT is lab-intensive, onerous, and needs expert skills. Besides, the manual results vary due to intraexpert and interexpert factors. Shan et al. (2021) [7] mention that “fully manual delineation that often takes hours” and the modern automatic diagnosis models based on artificial intelligence (AI) can only take seconds to minutes to get decisions, which now becomes a hot research field.

For example, Ye et al. (2011) [8] present a morphological classification (MC) method. The authors extract morphological features by registering cardiac MRI scans to a template. Later, deep learning (DL) rises as a new type of artificial intelligence (AI) technique and has shown its powerfulness in many academic and industrial fields. Within the field of

DL, convolutional neural network (CNN) is one standard DL algorithm that is particularly suitable for handling images. Giannakidis et al. (2016) [9] presented a multiscale three-dimensional CNN (3DCNN) for segmentation of the right ventricle. Tandon et al. (2021) [10] present a ventricular contouring CNN (VCCNN) algorithm.

The difference between this study to previous studies is that we simplify the problem to a binary-coded classification problem [11]; that is, given an input cardiovascular CT image, the AI model should have the ability to give a binary output, i.e., predict whether the subject is TOF or healthy. This simplification makes the AI model focus on the prediction task itself and does not need to generate human-understandable outputs (such as segmentation, contouring, etc.) in the light of the expectation to make our AI model more accurate. Furthermore, we propose a new stochastic pooling CNN (SCCNN) that uses a new pooling technique—stochastic pooling to improve the prediction performance. All in all, our contributes are fourfold:

- (a) Stochastic pooling is employed to replace traditional max-pooling
- (b) Structural optimization is carried out to fix the optimal structure
- (c) Multiple-way DA is introduced to increase the diversity of training images
- (d) Experiments by ten runs of 10-fold crossvalidation show that our method is better than three state-of-the-art approaches

The rest of this paper is structured as follows: Section 2 describes the dataset. Section 3 contains the rationale of methodology, including the preprocessing, stochastic pooling, structural optimization, multiple-way data augmentation, the implementation, Grad-CAM, and evaluation measures. Section 4 presents the experimental results and discussions. Section 5 concludes this paper.

## 2. Dataset

This study is a retrospective research, of which ethical approval is exempted. The imaging protocol is described below: Philips Brilliance 256 row spiral CT machine, KV: 80, MAS: 138, Layer Thickness 0.8mm, Lung Window (W: 1600 HU, L: -600 HU), Mediastinal Window (W: 750 HU, L: 90 HU), thin layer reconstruction according to the lesion display, layer thickness, and layer distance are both 0.8 mm mediastinal window images. Place the patient in a supine position, let the patient breathe deeply after holding in, and conventionally scan from the apex of the lung to the costal diaphragmatic angle. The resolutions of all images are 512 by 512 pixels. Data is available upon reasonable requests to corresponding authors.

We selected ten children with Tetralogy of Fallot who were admitted to Nanjing Children’s Hospital from March 2017 to March 2020. We then used a systematic random sampling method to select ten normal children from healthy

TABLE 1: Abbreviation list.

Abbreviation	Meaning
AP	Average pooling
AUC	Area under the curve
BN	Normalization
CNN	Convolutional neural network
CT	Computed tomography
DA	Data augmentation
DPD	Discrete probability distribution
FCL	Fully connected layer
FMI	Fowlkes–Mallows index
Grad-CAM	Gradient-weighted class activation mapping
GUI	Graphical user interface
HS	Histogram stretching
L2P	$l_2$ -norm pooling
MCC	Matthews correlation coefficient
MP	Max-pooling
MRI	Magnetic resonance imaging
MSD	Mean and standard deviation
PM	Probability map
ReLU	Rectified linear unit
RLV	Random location vector
ROC	Receiver operating characteristic
SC	Strided convolution
SP	Stochastic pooling
TOF	Tetralogy of Fallot

medical examiners within the same period of time. The Tetralogy of Fallot (TOF) observation group included three males and seven females, aged 4–22 months, with an average age of  $(8.90 \pm 5.47)$  months. Normal children in the control group included six males and four females, aged from 3 months to 24 months, with an average age of  $10.4 \pm 8.14$  months. Inclusion criteria for children with confirmed Tetralogy of Fallot are as follows:

- (1) CT suggests Tetralogy of Fallot
- (2) Surgery confirmed that the anatomical deformity of the heart is Tetralogy of Fallot

## 3. Methodology

*3.1. Preprocessing.* Table 1 lists the abbreviation list for the ease of reading. A five-step preprocessing was carried out on all the images to select the important slices, save storage, enhance contrast, remove unnecessary image regions, and reduce the image resolution.

First, four slices were chosen by radiologists using a slice-level selection method. For TOF patients, the slices showing the largest size and number of lesions were selected. For healthy control subjects, any level of the image can be selected. Now, we have in total 40 TOF images and 40 HC images.

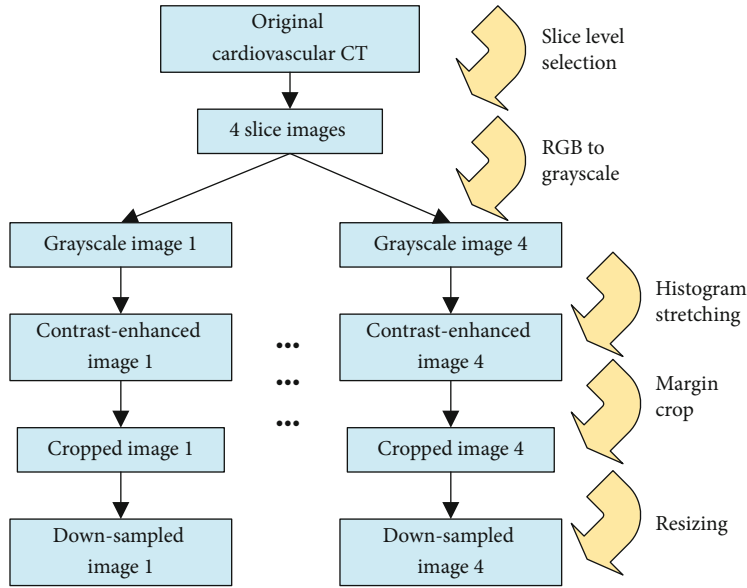


FIGURE 1: Diagram of preprocessing.

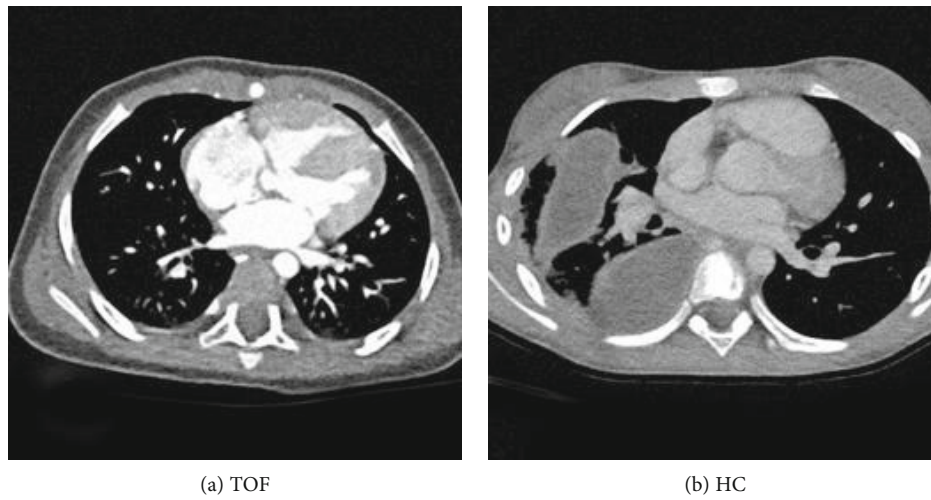


FIGURE 2: Illustration of our dataset.

Second, all the images are converted to grayscale images and stored in tiff format [12] using the compression lossless method. Third, histogram stretching (HS) was employed to enhance image contrast. Suppose the  $k$ th input and output of HS is  $x(k)$  and  $y(k)$ . HS can be formulated as

$$y(k) = \frac{x(k) - x_{\min}(k)}{x_{\max}(k) - x_{\min}(k)}, \quad (1)$$

where  $x_{\min}(k)$  and  $x_{\max}(k)$  stand for the minimum and maximum grayscale values in the input  $x(k)$ .

Fourth, cropping was done in order to eliminate the check-up bed at the bottom, the subject's two arms at bilateral sides, the rulers at the bottom and right side, and information (hospital, scanning protocol, subject's information, image head information, and labeling) at four corners.

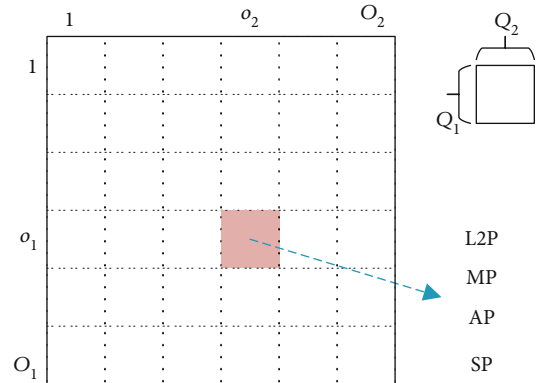


FIGURE 3: A diagram of block-wise pooling.

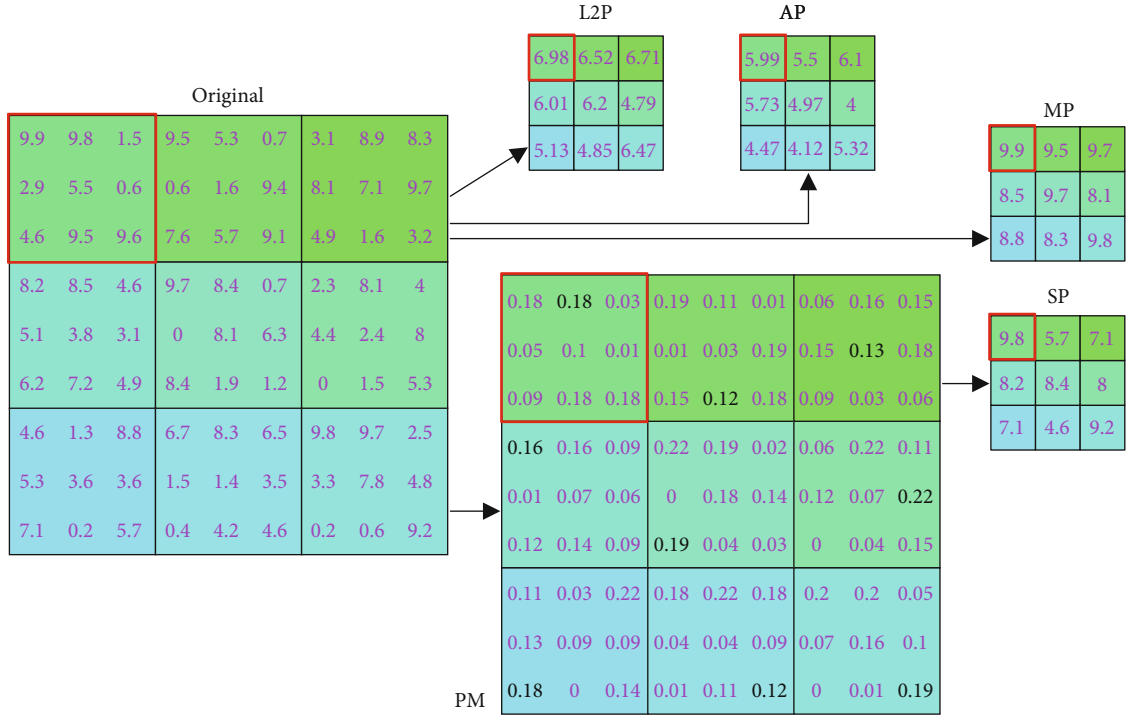
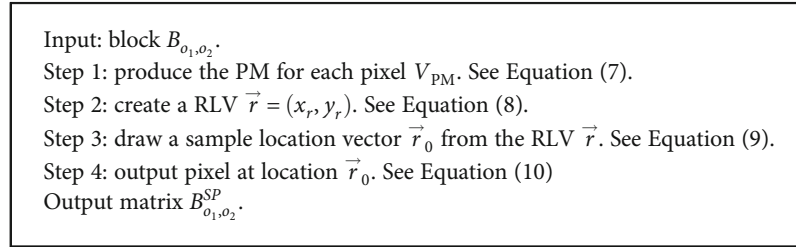


FIGURE 4: Comparison of four different pooling methods.



ALGORITHM 1: Pseudocode of SP.

Lastly, downscaling was performed to reduce each image to the size of  $[256 \times 256]$ . Figure 1 displays the diagram of our preprocessing procedure. Figures 2(a) and 2(b) shows two preprocessed examples of TOF and HC, respectively.

**3.2. Stochastic Pooling.** Pooling is an essential operation in standard convolutional neural networks (CNNs) [13]. Two types of pooling exist. One is max pooling (MP), and the other is average pooling (AP). The objective of pooling is to down-sample an input image or feature map (FM), reducing their dimensionality (width or height) and allowing for some assumption about the features to be made in each block.

Suppose we have an input image or FM, which can be split into  $O_1 \times O_2$  blocks, where every block has the extent of  $Q_1 \times Q_2$ . Currently, let us fix on the block  $B_{o_1, o_2}$  at  $o_1$ th row and  $o_2$ th column as shown as the red rectangle in Figure 3.

$$B_{o_1, o_2} = \{b(x, y), x = 1, \dots, Q_1, y = 1, \dots, Q_2\}, \quad (2)$$

TABLE 2: Structures of nine customized neural networks.

Configuration	No. of Conv layers	No. of FCLs
I	2	1
II	2	2
III	2	3
IV	3	1
V	3	2
VI	3	3
VII	4	1
VIII	4	2
IX	4	3

Bold means the best.

where  $1 \leq o_1 \leq O_1, 1 \leq o_2 \leq O_2$ ,  $b(x, y)$  means the pixel value at coordinate  $(x, y)$ .

The strided convolution (SC) goes over the input activation map with the strides that equals the size of the block  $(Q_1, Q_2)$ . The output of SC is

TABLE 3: Detailed structure of network of configuration V.

Layer	Parameters	FM
Input		256 × 256 × 1
Conv_1 (BN-ReLU)	32, 3 × 3/2	128 × 128 × 32
SP_1		64 × 64 × 32
Conv_2 (BN-ReLU)	64, 3 × 3/2	32 × 32 × 64
SP_2		16 × 16 × 64
Conv_3 (BN-ReLU)	128, 3 × 3	16 × 16 × 128
SP_3		8 × 8 × 128
Flatten		8192
FCL_1	100 × 8192, 100 × 1	100
FCL_2	2 × 100, 2 × 1	2
Output		

$$B_{o_1, o_2}^{SC} = b(1, 1). \quad (3)$$

The  $l_2$ -norm pooling (L2P), average pooling (AP) [14], and max pooling (MP) [15] produce the  $l_2$ -norm, average, and maximum values within the block  $B_{m_1, m_2}$ , respectively. Their formula can be written as below:

$$B_{o_1, o_2}^{L2P} = \sqrt{\frac{\sum_{x=1}^{Q_1} \sum_{y=1}^{Q_2} b^2(x, y)}{Q_1 \times Q_2}}, \quad (4)$$

$$\begin{cases} P[\vec{r} = (1, 1)] = V_{PM}(1, 1) & P[\vec{r} = (1, 2)] = V_{PM}(1, 2) & \cdots & P[\vec{r} = (1, Q_2)] = V_{PM}(1, Q_2), \\ P[\vec{r} = (2, 1)] = V_{PM}(2, 1) & P[\vec{r} = (2, 2)] = V_{PM}(2, 2) & \cdots & P[\vec{r} = (2, Q_2)] = V_{PM}(2, Q_2), \\ \cdots & \cdots & \cdots & \cdots \\ P[\vec{r} = (Q_1, 1)] = V_{PM}(Q_1, 1) & P[\vec{r} = (Q_1, 2)] = V_{PM}(Q_1, 2) & \cdots & P[\vec{r} = (Q_1, Q_2)] = V_{PM}(Q_1, Q_2), \end{cases} \quad (8)$$

where  $P$  represents the probability. Shortly speaking,  $P[\vec{r} = (x, y)] = V_{PM}(x, y)$ ,  $\forall 1 \leq x \leq Q_1 \& 1 \leq y \leq Q_2$  or  $\vec{r} \sim V_{PM}$ , namely, the distribution of RLV  $\vec{r}$  has the DPD as  $V_{PM}$ .

*Step 3.* A sample location vector  $\vec{r}_0$  is drawn from the RLV  $\vec{r}$ , and we have

$$\vec{r}_0 = (x_{r_0}, y_{r_0}). \quad (9)$$

*Step 4.* SP outputs the pixel at the location  $\vec{r}_0$ , namely,

$$B_{o_1, o_2}^{SP} = b(x_{r_0}, y_{r_0}). \quad (10)$$

$$B_{o_1, o_2}^{AP} = \frac{1}{Q_1 \times Q_2} \sum_{x=1}^{Q_1} \sum_{y=1}^{Q_2} b(x, y), \quad (5)$$

$$B_{o_1, o_2}^{MP} = \max_{x=1}^{Q_1} \max_{y=1}^{Q_2} b(x, y). \quad (6)$$

Nevertheless, the AP outputs the average, downscaling the greatest value, where the important features may lie. In contrast, MP stores the greatest value but deteriorates the overfitting obstacle. In order to solve the above concerns, stochastic pooling (SP) [15] is introduced to provide a resolution to the drawbacks of AP and MP. SP is a four-step process.

*Step 1.* It produces the probability map (PM)  $V_{PM}$  for each pixel in the block  $B_{o_1, o_2}$ .

$$\begin{cases} V_{PM}(x, y) = \frac{b(x, y)}{\sum_{x=1}^{Q_1} \sum_{y=1}^{Q_2} b(x, y)}, \\ \text{s.t. } \sum_{x=1}^{Q_1} \sum_{y=1}^{Q_2} V_{PM}(x, y) = 1, \end{cases} \quad (7)$$

where  $V_{PM}(x, y)$  stands for the PM value at pixel  $(x, y)$ .

*Step 2.* It creates a random location vector (RLV)  $\vec{r} = (x_r, y_r)$  that takes the discrete probability distribution (DPD) as

Figure 4 shows a realistic example of four different pooling methods. Algorithm 1 presents the pseudocode of SP. Take the  $3 \times 3$  block  $B_{1,1}$  (The red rectangle in Figure 4) as an example, L2P generates the output as 6.98. AP and MP present 5.99 and 9.9, respectively. Meanwhile, SP first generates PM matrix

$$V_{PM} = \begin{bmatrix} 0.18 & 0.18 & 0.03 \\ 0.05 & 0.1 & 0.01 \\ 0.09 & 0.18 & 0.18 \end{bmatrix}, \quad (11)$$

and a sample location vector is drawn as  $\vec{r}_0 = (1, 2)$ . Therefore, the output of SP is  $B_{1,1}^{SP} = b(1, 2) = 9.8$ .

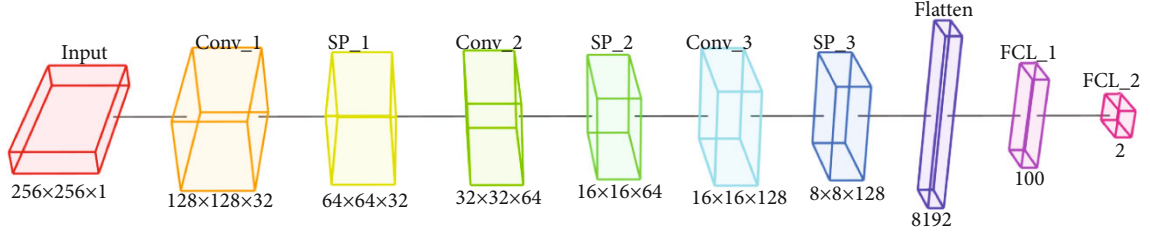


FIGURE 5: FM plot.

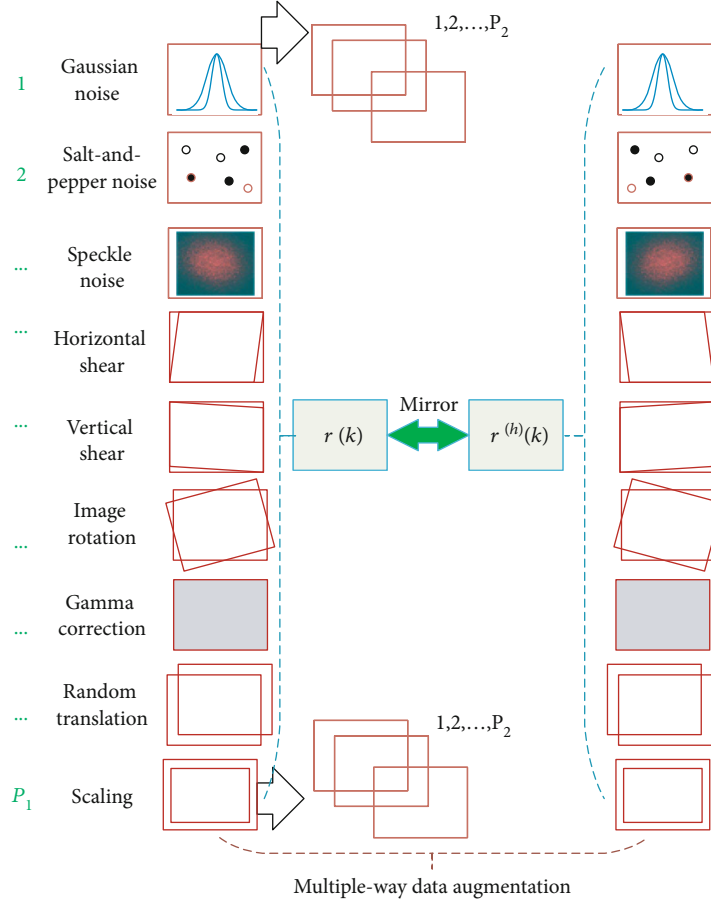


FIGURE 6: Diagram of multiple-way data augmentation.

Input: import raw preprocessed  $k$ th training image  $r(k)$ .

$P_1$  geometric or photometric or noise-injection DA transforms  $Z_p$  are utilized on  $r(k)$ .

Step 1: we obtain  $Z_p[r(k)]$ ,  $p = 1, \dots, P_1$ . See Equation (12)

Each enhanced dataset contains  $P_2$  new images. See Equation (13).

Step 2: a horizontal mirror image is produced as  $r^{(h)}(k) = \beta_1[r(k)]$ . See Equation (14).

Step 3:  $M_1$ -way data augmentation methods are carried out on  $r^{(h)}(k)$ ,

We obtain  $Z_p[r^{(h)}(k)]$ ,  $p = 1, \dots, P_1$ . See Equation (15).

Step 4:  $r(k)$ ,  $r^{(h)}(k)$ ,  $Z_p[r(k)]$ ,  $p = 1, \dots, P_1$ , and  $Z_p[r^{(h)}(k)]$ ,  $p = 1, \dots, P_1$  are merged via  $\beta_2$ . See Equation (16).

Output A new dataset  $G(k)$  is produced. The number of images is  $P_3 = 2 \times P_1 \times P_2 + 2$ . See Equation (17).

ALGORITHM 2: Pseudocode of proposed 18-way DA on  $k$ th training image

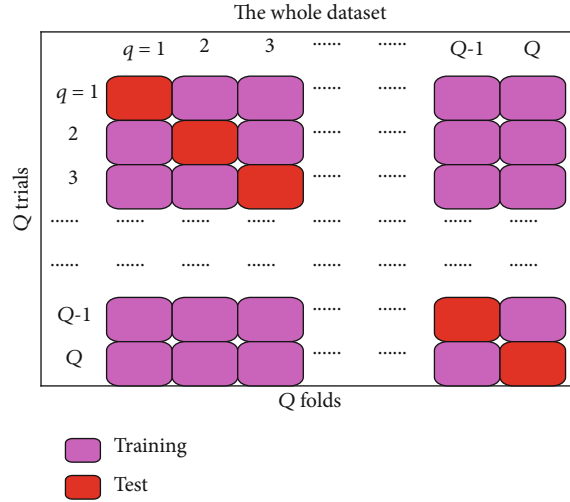


FIGURE 7: Q-fold crossvalidation.

TABLE 4: Meanings in measures.

Abbreviation	Full form	Symbol	Meaning
$P$	Positive		TOF
$N$	Negative		HC
TP	True positive	$g(1, 1)$	TOF images are classified correctly.
FP	False positive	$g(2, 1)$	HC images are wrongly classified as TOF.
TN	True negative	$g(2, 2)$	HC images are classified correctly.
FN	False negative	$g(1, 2)$	TOF images are wrongly classified as HC.

TABLE 5: Statistical analysis of SOSPCNN model.

Run	Sen	Spc	Prc	Acc	F1	MCC	FMI
1	95.00	92.50	92.68	93.75	93.83	87.53	93.83
2	92.50	90.00	90.24	91.25	91.36	82.53	91.36
3	95.00	92.50	92.68	93.75	93.83	87.53	93.83
4	90.00	92.50	92.31	91.25	91.14	82.53	91.15
5	90.00	95.00	94.74	92.50	92.31	85.11	92.34
6	90.00	97.50	97.30	93.75	93.51	87.75	93.58
7	92.50	95.00	94.87	93.75	93.67	87.53	93.68
8	92.50	90.00	90.24	91.25	91.36	82.53	91.36
9	90.00	92.50	92.31	91.25	91.14	82.53	91.15
10	95.00	90.00	90.48	92.50	92.68	85.11	92.71
MSD	$92.25 \pm 2.19$	$92.75 \pm 2.49$	$92.79 \pm 2.29$	$92.50 \pm 1.18$	$92.48 \pm 1.17$	$85.06 \pm 2.38$	$92.50 \pm 1.17$

3.3. *Structural Optimization.* How to obtain the best network structure [16]? We try to design nine different configurations in this study. Their hyperparameters of structures are listed in Table 2. Two hyperparameters are considered in this study: (i) the number of Conv layers and (ii) the number of fully connected layers (FCLs). Those two types of layers are common layers in standard CNN, so we will not introduce them due to the page limit.

In the following experiment, we will observe that configuration V, a five-layer customized neural network, gives the

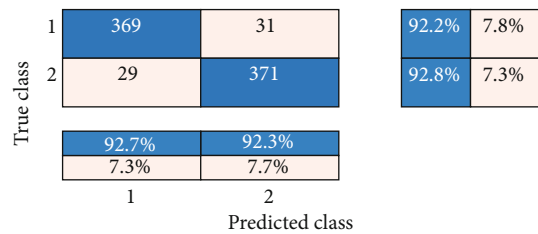


FIGURE 8: Confusion matrix of  $10 \times 10$ -fold crossvalidation (Here, classes 1 and 2 stand for ToF and HC, respectively).

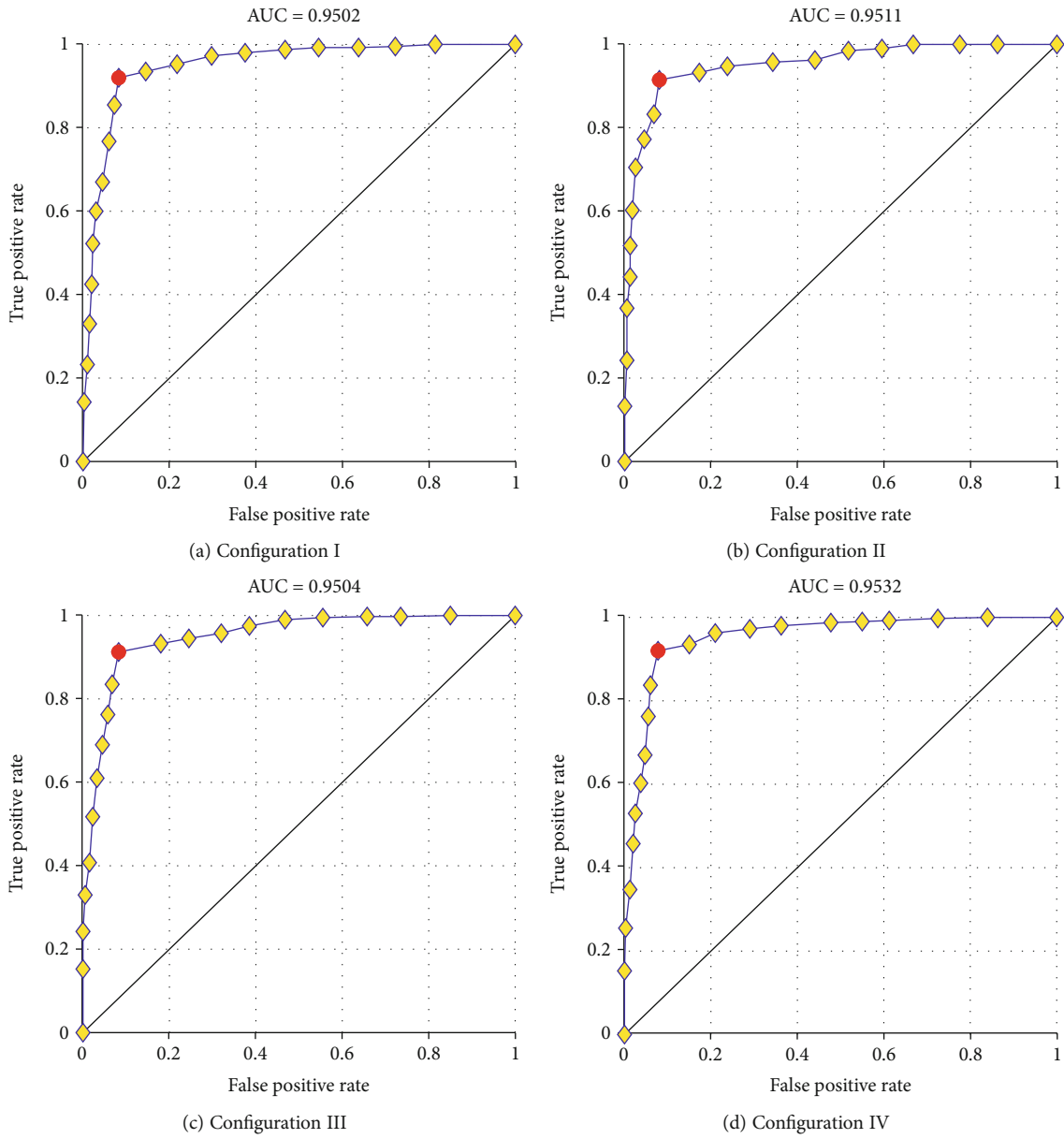
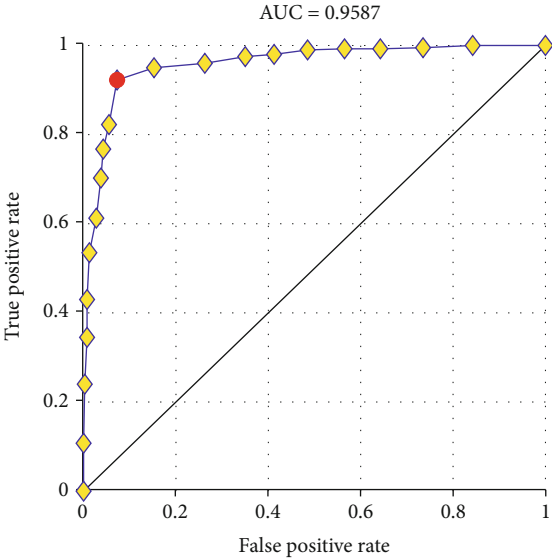
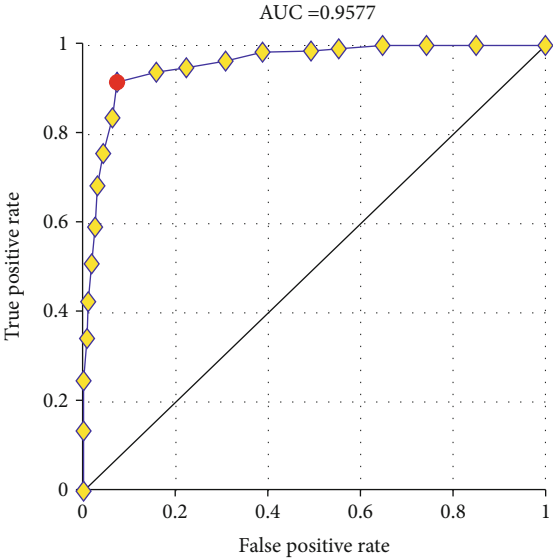


FIGURE 9: Continued.

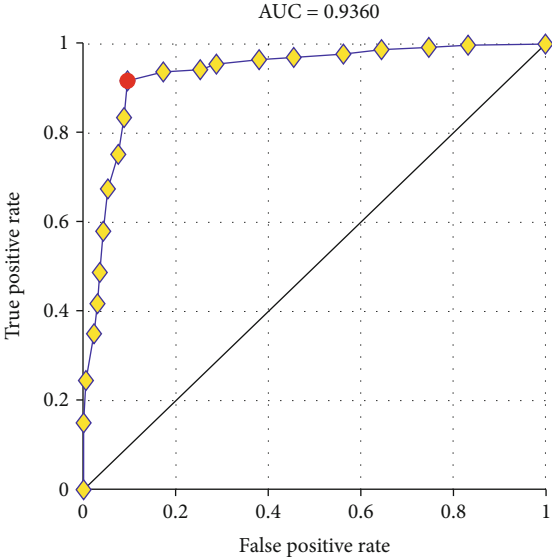




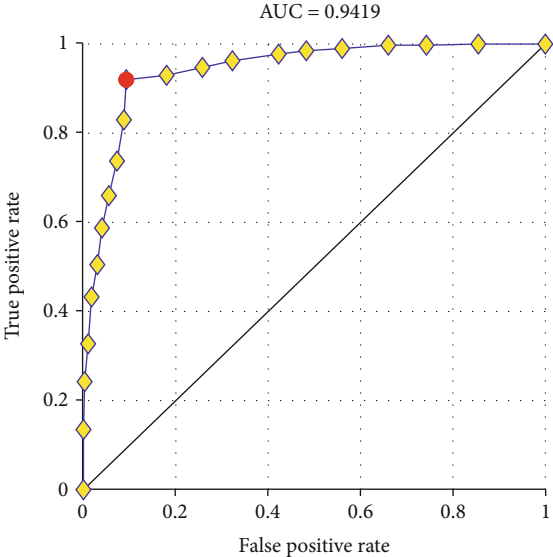
(e) Configuration V



(f) Configuration VI

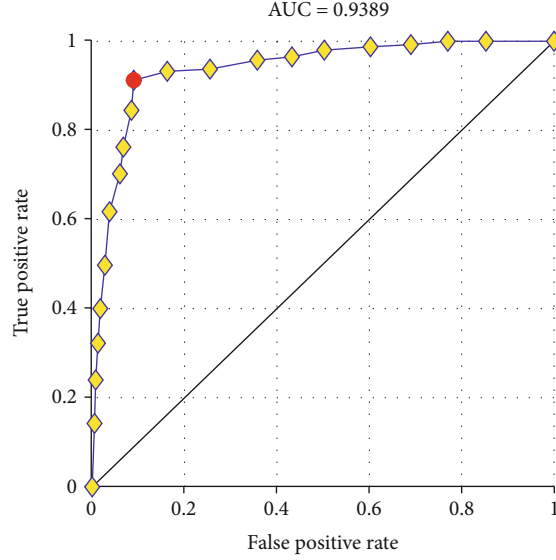


(g) Configuration VII



(h) Configuration VIII

FIGURE 9: Continued.



(i) Configuration IX

FIGURE 9: Comparison of nine configurations.

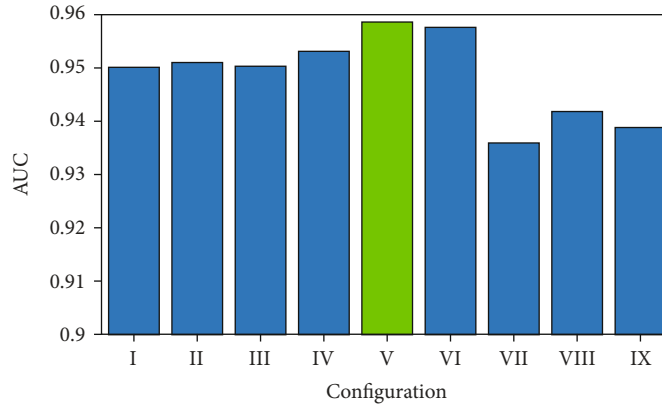


FIGURE 10: Bar plot of AUC against nine configurations.

best performance. Here, we briefly give its detailed structure in Table 3. The input is of size  $256 \times 256 \times 1$ . The first Conv layer (Conv\_1) is associated with the batch normalization (BN) layer and rectified linear unit (ReLU) activation. The parameters of Conv\_1 are 32 kernels with sizes of  $3 \times 3$  and stride of 2. Afterward, the first SP (SP\_1) reduce the FM from  $128 \times 128 \times 32$  to  $64 \times 64 \times 32$ .

After three Conv layers and three SP layers, the size of FM is  $8 \times 8 \times 128$ . It is then flattened to a vector of 8192 neurons. With two FCLs of 100 and 2 hidden neurons, the neural network finally outputs whether TOF or HC. All in all, our model is termed structurally optimized stochastic pooling convolutional neural network (SOSPCNN). The FM plot is portrayed in Figure 5.

**3.4. Multiple-Way Data Augmentation.** The relatively small dataset ( $40+40=80$  images) may bring the overfitting problem. To avoid overfitting, data augmentation (DA) [17] is a powerful tool because it can generate synthetic images on the training set [18]. Zhu (2021) [19] presented an 18-way

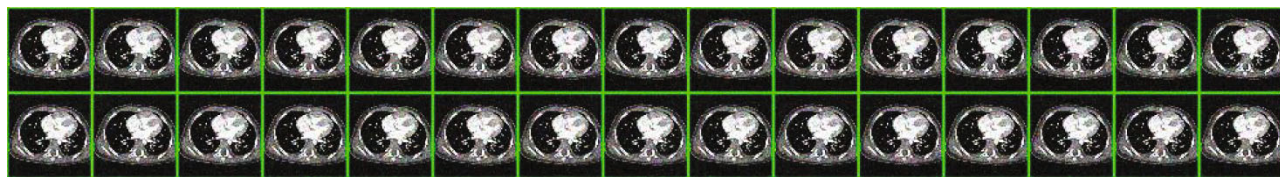
DA method and proved this 18-way DA works better than the traditional DA approach. Its diagram is shown in Figure 6. The difference of DA and MDA is that (i) MDA uses a combination of different DA methods on training set; (ii) MDA is modular design. The users are easy to add or remove particular DA methods from a MDA.

Suppose we have the raw training image  $r(k)$ , where  $k$  represents the image index. First,  $P_1$  different DA methods displayed in Figure 6 are applied to  $r(k)$ . Let  $Z_p, p = 1, \dots, P_1$  be each DA operation, we get  $P_1$  augmented datasets on raw image  $r(k)$  as

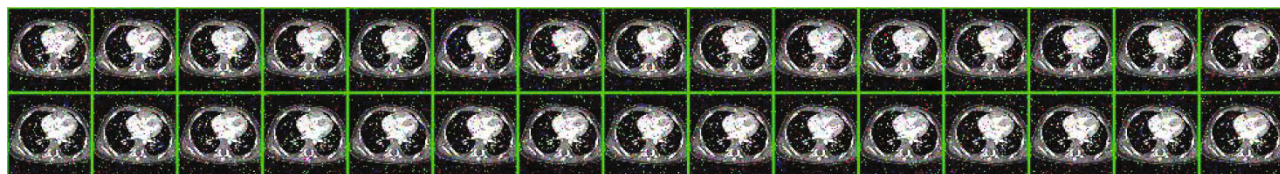
$$Z_p[r(k)], p = 1, \dots, P_1. \quad (12)$$

Let  $P_2$  stands for the size of generated new images for each DA method, thus,

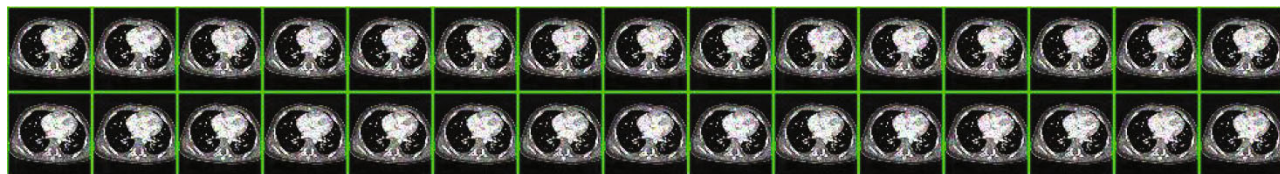
$$|Z_p[r(k)]| = P_2. \quad (13)$$



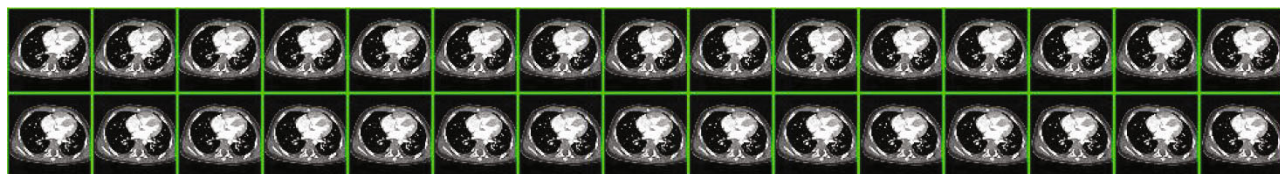
(a) Gaussian noise



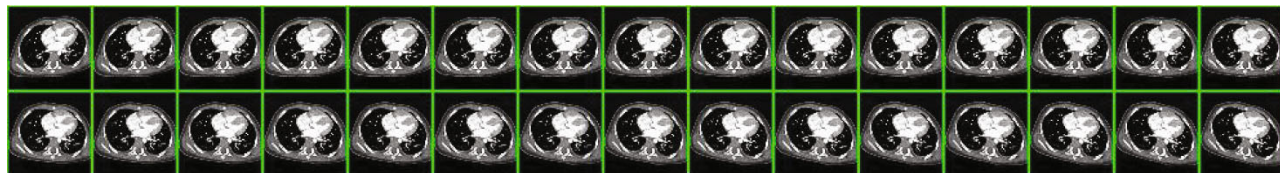
(b) Salt-and-pepper noise



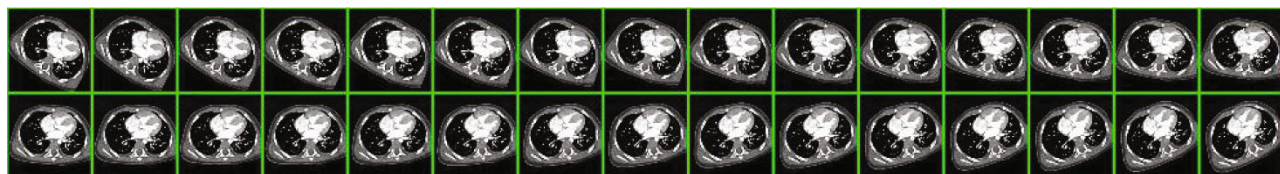
(c) Speckle noise



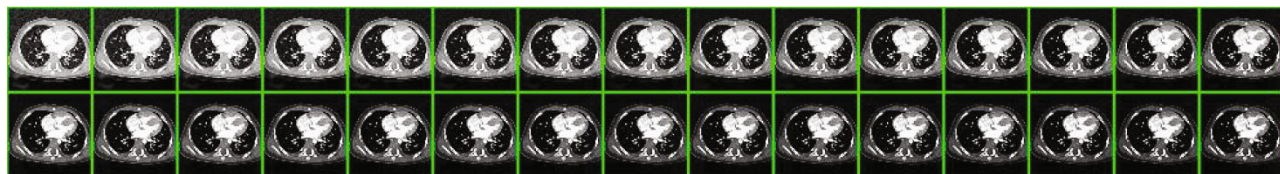
(d) Horizontal shear



(e) Vertical shear



(f) Rotation



(g) Gamma correction

FIGURE 11: Continued.

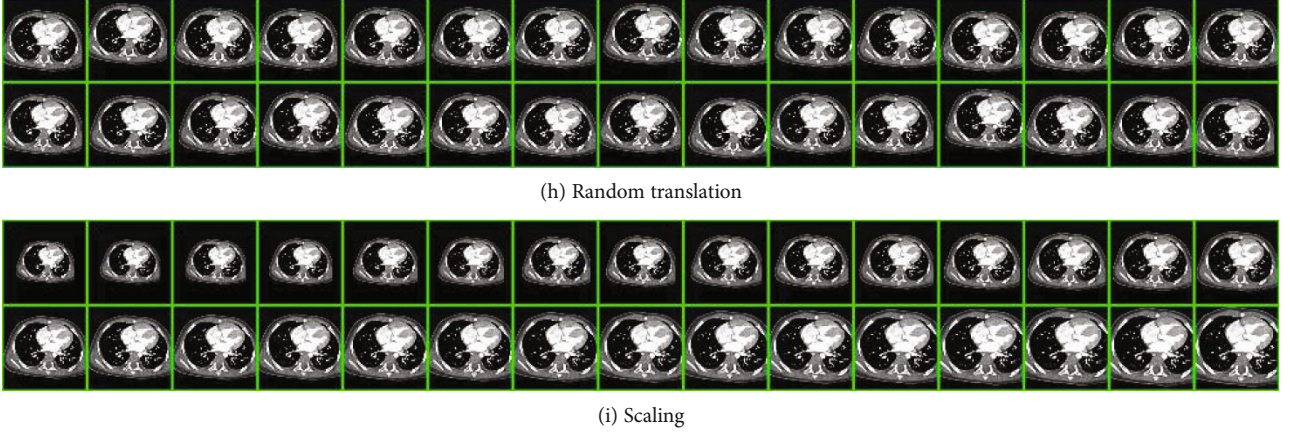


FIGURE 11: Illustration of multiple-way data augmentation.

Second, horizontal mirrored image is produced by

$$r^{(h)}(k) = \beta_1[r(k)], \quad (14)$$

where  $\beta_1$  means horizontal mirror function.

Third, all  $P_1$  different DA methods are carried out on the mirrored image  $r^{(h)}(k)$  and produce  $P_1$  new datasets as

$$\begin{cases} \mathbf{Z}_p[r^{(h)}(k)], & p = 1, \dots, P_1, \\ |\mathbf{Z}_p[r^{(h)}(k)]| = P_2, & p = 1, \dots, P_1. \end{cases} \quad (15)$$

Fourth, the raw image  $r(k)$ , the mirrored image  $r^{(h)}(k)$ , all  $P_1$ -way results of raw image  $\mathbf{Z}_p[r(k)]$ , and all  $P_1$ -way DA results of horizontal mirrored image  $\mathbf{Z}_p[r^{(h)}(k)]$  are combined. The final generated dataset from  $r(k)$  is defined as  $\mathbf{G}(k)$ :

$$r(k) \mapsto \mathbf{G}(k) = \beta_2 \left\{ \begin{array}{cc} r(k) & r^{(h)}(k) \\ \underbrace{\mathbf{Z}_1[r(k)]}_{P_2} & \underbrace{\mathbf{Z}_1[r^{(h)}(k)]}_{P_2} \\ \dots & \dots \\ \underbrace{\mathbf{Z}_{P_1}[r(k)]}_{P_2} & \underbrace{\mathbf{Z}_{P_1}[r^{(h)}(k)]}_{P_2} \end{array} \right\}, \quad (16)$$

where  $\beta_2$  stands for the concatenation function. Let augmentation factor be  $P_3$ , which means the number of images in  $\mathbf{G}(k)$ , we obtain

$$P_3 = \frac{|\mathbf{G}(k)|}{|r(k)|} = \frac{(1 + P_1 \times P_2) \times 2}{1} = 2 \times P_1 \times P_2 + 2. \quad (17)$$

Algorithm 2 recaps the pseudocode of this 18-way DA. We set  $P_1 = 9$  to achieve an 18-way DA. We also set  $P_2 = 30$ , thus  $P_3 = 542$ , indicating each raw training image will generate 542 images, which include the raw image  $r(k)$  itself.

**3.5. Implementation and Grad-CAM.**  $Q$ -fold crossvalidation [20] is employed. The whole dataset is divided into  $Q$  folds (see Figure 7). At  $q$ th trial,  $1 \leq q \leq Q$ , the  $q$ th fold is picked up as the test, and the rest  $Q - 1$  folds:  $[1, \dots, q - 1, q + 1, \dots, Q]$  are chosen as training set [21]. In this study, we set  $Q = 10$ , namely, a 10-fold cross validation. Furthermore, we run the 10-fold crossvalidation 10 times, i.e.,  $10 \times 10$ -fold crossvalidation.

Gradient-weighted class activation mapping (Grad-CAM) [22] is employed to explain how our model makes its decision in classification. Grad-CAM utilizes the gradient of the classification score with respect to the convolutional features determined by the network to understand which parts of the image are most important for classification. Grad-CAM is a generalization of the class activation mapping (CAM) method [23] to a broader range of CNN models since the original CAM relies on a fully convolutional neural network structure. The output of SP\_3 (see Table 3) is used as the feature layer for Grad-CAM.

Mathematically, suppose our classification network is with output  $y^c$ , standing for the score for class  $c$ . We would like to compute the Grad-CAM map for a layer with  $k$  feature maps  $A_{i,j}^k$ , where  $(i, j)$  stands for the indexes of pixels. We can obtain the neural importance weight as

$$\alpha_k^c = \frac{1}{N} \sum_i \sum_j \frac{\partial y^c}{\partial A_{i,j}^k}, \quad (18)$$

where  $N$  stands for the total number of pixels in the feature map. The Grad-CAM is a weighted combination of the feature maps with a ReLU as

$$M = \text{ReLU} \left( \sum_k \alpha_k^c A^k \right). \quad (19)$$

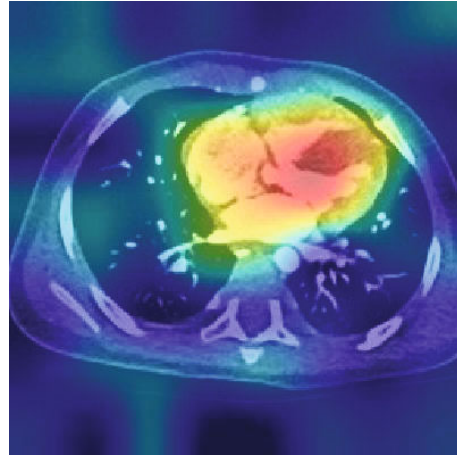
The Grad-CAM map  $M$  is then upsampled to the size of input data.

TABLE 6: Statistical analysis without multiple-way data augmentation.

Run	Sen	Spc	Prc	Acc	F1	MCC	FMI
1	87.50	90.00	89.74	88.75	88.61	77.52	88.61
2	90.00	87.50	87.80	88.75	88.89	77.52	88.90
3	87.50	87.50	87.50	87.50	87.50	75.00	87.50
4	90.00	87.50	87.80	88.75	88.89	77.52	88.90
5	85.00	90.00	89.47	87.50	87.18	75.09	87.21
6	85.00	90.00	89.47	87.50	87.18	75.09	87.21
7	82.50	92.50	91.67	87.50	86.84	75.38	86.96
8	87.50	87.50	87.50	87.50	87.50	75.00	87.50
9	85.00	90.00	89.47	87.50	87.18	75.09	87.21
10	82.50	92.50	91.67	87.50	86.84	75.38	86.96
MSD	86.25 ± 2.70	89.50 ± 1.97	89.21 ± 1.58	87.87 ± 0.60	87.66 ± 0.82	75.86 ± 1.16	87.70 ± 0.79



(a) Manually delineated



(b) Heatmap

FIGURE 12: Heatmap of one TOF image.

3.6. *Measures.* The confusion matrix of 10 runs of 10-fold crossvalidation is supposed to be

$$G = \begin{bmatrix} g(1,1) & g(1,2) \\ g(2,1) & g(2,2) \end{bmatrix} = \begin{bmatrix} TP & FN \\ FP & TN \end{bmatrix}. \quad (20)$$

Note  $FN = FP = 0$  for a perfect classification. The meaning of  $P$ ,  $N$ ,  $TP$ ,  $FP$ ,  $TN$ , and  $FN$  are itemized in Table 4.

Nine measures are used: sensitivity, specificity, precision, accuracy, F1 score, Matthews correlation coefficient (MCC), Fowlkes–Mallows index (FMI), receiver operating characteristic (ROC), and area under the curve (AUC).

The first four measures are defined as

$$\begin{cases} \text{Sen} = \frac{g(1,1)}{g(1,1) + g(1,2)} & \text{Spc} = \frac{g(2,2)}{g(2,2) + g(2,1)}, \\ \text{Prc} = \frac{g(1,1)}{g(1,1) + g(2,1)} & \text{Acc} = \frac{g(1,1) + g(2,2)}{g(1,1) + g(2,2) + g(1,2) + g(2,1)}. \end{cases} \quad (21)$$

F1, MCC [24], and FMI [25] are defined as

$$\begin{cases} F_1 = 2 \times \frac{\text{Sen} \times \text{Prc}}{\text{Sen} + \text{Prc}} = \frac{2 \times g(1,1)}{2 \times g(1,1) + g(1,2) + g(2,1)}, \\ \text{MCC} = \frac{g(1,1) \times g(2,2) - g(2,1) \times g(1,2)}{\sqrt{[g(1,1) + g(2,1)] \times [g(1,1) + g(1,2)] \times [g(2,2) + g(2,1)] \times [g(2,2) + g(1,2)]}}, \\ \text{FMI} = \sqrt{\text{Sen} \times \text{Prc}} = \sqrt{\frac{g(1,1)}{g(1,1) + g(1,2)} \times \frac{g(1,1)}{g(1,1) + g(2,1)}}. \end{cases} \quad (22)$$

The above measures are calculated in the mean and standard deviation (MSD) format. Furthermore, ROC is a curve to measure a binary classifier with varying discrimination thresholds [26]. The ROC curve is created by plotting the sensitivity against 1-specificity. The AUC is calculated based on the ROC curve [27].

## 4. Experimental Results

4.1. *Statistical Analysis.* The result of the SOSPCNN model using configuration V is itemized in Table 5. The model arrives at a performance with a sensitivity of  $92.25 \pm 2.19$ , a specificity of  $92.75 \pm 2.49$ , a precision of  $92.79 \pm 2.29$ , an

TABLE 7: Comparison with state-of-the-art approaches.

Approach	Sen	Spc	Prc	Acc	F1	MCC	FMI
MC [8]	86.25 ± 3.58	80.75 ± 3.55	81.88 ± 2.32	83.50 ± 0.79	83.92 ± 0.97	67.25 ± 1.56	84.00 ± 0.98
3DCNN [9]	91.00 ± 3.16	89.50 ± 3.29	89.77 ± 2.70	90.25 ± 1.42	90.32 ± 1.42	80.63 ± 2.79	90.35 ± 1.41
VCCNN [10]	90.75 ± 1.69	90.00 ± 2.36	90.14 ± 1.95	90.38 ± 0.84	90.41 ± 0.78	80.80 ± 1.64	90.43 ± 0.76
SOSPCNN (ours)	92.25 ± 2.19	92.75 ± 2.49	92.79 ± 2.29	92.50 ± 1.18	92.48 ± 1.17	85.06 ± 2.38	92.50 ± 1.17

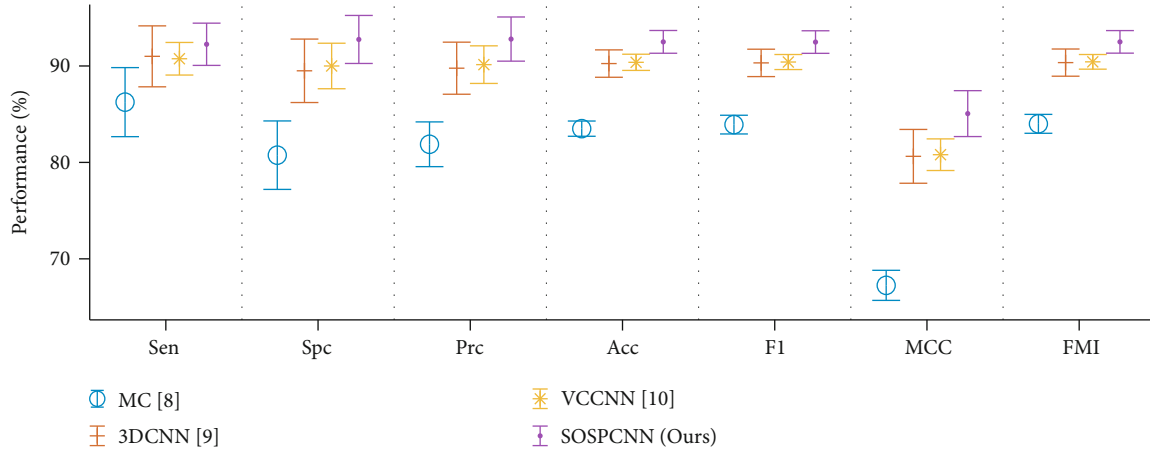


FIGURE 13: Error bar comparison.

accuracy of  $92.50 \pm 1.18$ , an F1 score of  $92.48 \pm 1.17$ , an MCC of  $85.06 \pm 2.38$ , and an FMI of  $92.50 \pm 1.17$ .

Figure 8 shows the confusion matrix of  $10 \times 10$ -fold crossvalidation, where we can see the TP = 369, FN = 31, TN = 371, and FP = 29, indicating 31 TOF are wrongly classified as HC while 29 HC are misclassified to TOF. Hence, the sensitivity is  $369/(369 + 31) = 92.25\%$ , and specificity is  $371/(29 + 371) = 92.75\%$ .

**4.2. Configuration Comparison.** We compare nine configurations (see Table 2). The validation is the same as previous experiment. Due to the page limit, the detailed statistical analysis is not shown. The ROC and AUC values are displayed in Figure 9. The AUC values of nine networks with different configurations are: 0.9502, 0.9511, 0.9504, 0.9532, 0.9587, 0.9577, 0.9360, 0.9419, and 0.9389 (as shown in Figure 10). We can observe from Figure 10 that the best network is with configuration V, whose structure is shown in Table 3.

**4.3. Effect of Multiple-Way Data Augmentation.** Figure 11 shows the multiple-way DA results if we take Figure 2(a) as the raw training examples. Due to the page limit, the multiple-way DA results on the horizontally mirrored image are not displayed. As we can see from Figure 11, multiple-way DA increases the diversity of the training images.

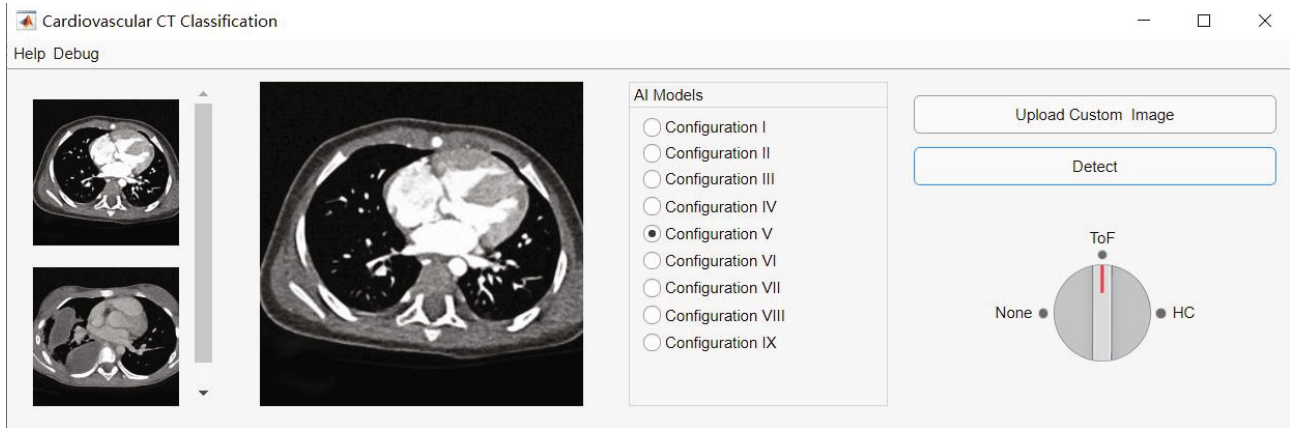
If we remove the multiple-way data augmentation from our model, the performances are decreased, as shown in Table 6, where MSD stands for mean and standard deviation. Comparing Table 5 with Table 6, we can observe multiple-way DA is efficient in improving the classification performance. The reason is that it helps our model resist overfitting by enhancing the diversity of the training set.

**4.4. Explainability.** Figure 12 shows the manual delineation and the heat map of Figure 2(a) via Grad-CAM described in Section 3.5. The manual delineation showed the radiologist make decisions on “TOF” diagnosis based on all the areas of the abnormal heart, while the heat map shows the proposed SOSPCNN model also puts more focus on the heart region other than the surrounding tissues and background areas.

**4.5. Comparison with State-of-the-Art Approaches.** We compare the proposed SOSPCNN model with three other approaches: MC [8], 3DCNN [9], and VCCNN [10]. The results are shown in Table 7. Note that some comparison methods are not suitable for our dataset, so we modify them to adapt to our dataset.

The error bar comparison is drawn in Figure 13, which clearly shows that the proposed SOSPCNN outperforms all three comparative approaches. The reason is three folds: (i) we use stochastic pooling to replace traditional max-pooling; (ii) we use structural optimization to determine the optimal structure of our SOSPCNN model; (iii) multiple-way DA is included to increase the diversity of training images. In the future, more advanced techniques [28–30] will be tested and integrated into our model.

**4.6. Desktop and Web Apps.** MATLAB app designer is used to create a professional application for both desktop and web. The input to this web app is any cardiovascular CT image, and the aforementioned SOSPCNN model is included in our app. Figure 14(a) displays the graphical user interface (GUI) of the standalone desktop app. The users can upload their custom images, and the software can show the results by turning the knob into the correct texts: TOF, HC, or none.



(a) Desktop app



(b) Web app

FIGURE 14: GUI of developed apps.

Figure 14(b) shows the GUI of the web app that is accessed through a “Google Chrome” web browser. The web app is based on a client-server modeled structure [31], i.e., the user is provided services through an off-site server hosted by a third-party cloud service, Microsoft Azure in our study. Our developed online web app can assist hospital clinicians in making decisions remotely and effectively.

## 5. Conclusion

This paper proposes a web app for TOF recognition. Our proposed model is termed structurally optimized stochastic pooling convolutional neural network (SOSPCNN) with explainable property achieved by Grad-CAM. The results by ten runs of 10-fold crossvalidation show that this SOSPCNN model yields a sensitivity of  $92.25 \pm 2.19$ , a specificity of  $92.75 \pm 2.49$ , a precision of  $92.79 \pm 2.29$ , an accuracy of  $92.50 \pm 1.18$ , an F1 score of  $92.48 \pm 1.17$ , an MCC of  $85.06 \pm 2.38$ , an FMI of  $92.50 \pm 1.17$ , and an AUC of 0.9587. Further, we develop both desktop and web apps to realize this SOSPCNN model.

The shortcomings of our method are as follows: (i) our model is trained on a small dataset; (ii) our model does not go through strict medical verification; (iii) our model only considers TOF and HC.

Therefore, we shall attempt to solve the above three weak points in the future. We shall try to collect more TOF and HC cardiovascular CT images. We shall invite clinicians to use our web app and return feedbacks so that we can continue to improve our model. We shall try to collect data of other heart diseases, so make our model can identify more types of diseases.

## Data Availability

Data is available upon reasonable requests to corresponding authors.

## Conflicts of Interest

The authors declare that they have no conflicts of interest to report regarding the present study.

## Authors' Contributions

Shui-Hua Wang and Kaihong Wu contributed equally to this work.

## Acknowledgments

This paper is partially supported by the Royal Society International Exchanges Cost Share Award, UK (RP202G0230); Medical Research Council Confidence in Concept Award, UK (MC\_PC\_17171); Hope Foundation for Cancer Research, UK (RM60G0680); British Heart Foundation Accelerator Award, UK; Sino-UK Industrial Fund, UK (RP202G0289); and Global Challenges Research Fund (GCRF), UK (P202PF11).

## References

- [1] D. Carli, A. Moroni, A. Zonta et al., "Atypical microdeletion 22q11.2 in a patient with tetralogy of Fallot," *Journal of Genetics*, vol. 100, no. 1, pp. 1–4, 2021.
- [2] M. Ghaderian, A. Ahmadi, M. R. Sabri et al., "Clinical outcome of right ventricular outflow tract stenting versus Blalock-Taussig shunt in Tetralogy of Fallot: a systematic review and meta-analysis," *Current Problems in Cardiology*, vol. 46, no. 3, article 100643, 2021.
- [3] M. Uecker, C. Petersen, C. Dingemann, C. Fortmann, B. M. Ure, and J. Dingemann, "Gravitational autoreposition for staged closure of omphaloceles," *European Journal of Pediatric Surgery*, vol. 30, no. 1, pp. 45–50, 2020.
- [4] E. Cambronero-Cortinas, P. Moratalla-Haro, A. E. González-García et al., "Predictors of atrial tachyarrhythmias in adults with congenital heart disease," *Kardiologia Polska*, vol. 78, no. 12, pp. 1262–1270, 2020.
- [5] T. Ashraf, F. Farooq, A. S. Muhammad et al., "Coronary artery anomalies in Tetralogy of Fallot patients undergoing CT angiography at a tertiary care hospital," *Cureus*, vol. 12, no. 9, article e10723, 2020.
- [6] M. Engbersen, M. Versleijen, D. Lambregts, R. Beets-Tan, M. Tesselaaar, and L. Max, "Comparison of whole-body MRI and 68Ga-DOTATATE PET-CT findings in patients with suspected peritoneal metastases from neuroendocrine tumors," *Journal of Neuroendocrinology*, vol. 33, pp. 120–120, 2021.
- [7] F. Shan, Y. Gao, J. Wang et al., "Abnormal lung quantification in chest CT images of COVID-19 patients with deep learning and its application to severity prediction," *Medical Physics*, vol. 48, no. 4, pp. 1633–1645, 2021.
- [8] D. H. Ye, H. Litt, C. Davatzikos, and K. M. Pohl, "Morphological classification: application to cardiac MRI of Tetralogy of Fallot," in *Functional Imaging and Modeling of the Heart*, D. N. Metaxas and L. Axel, Eds., pp. 180–187, Springer-Verlag Berlin, Berlin, 2011.
- [9] A. Giannakidis, K. Kamnitsas, V. Spadotto et al., "Fast fully automatic segmentation of the severely abnormal human right ventricle from cardiovascular magnetic resonance images using a multi-scale 3D convolutional neural network," in *2016 12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, pp. 42–46, New York, 2016.
- [10] A. Tandon, N. Mohan, C. Jensen et al., "Retraining convolutional neural networks for specialized cardiovascular imaging tasks: lessons from tetralogy of Fallot," *Pediatric Cardiology*, vol. 42, no. 3, pp. 578–589, 2021.
- [11] M. Klimo, P. Lukáč, and P. Tarábek, "Deep neural networks classification via binary error-detecting output codes," *Applied Sciences*, vol. 11, no. 8, article 3563, 2021.
- [12] V. S. Alfio, D. Costantino, and M. Pepe, "Influence of image TIFF format and JPEG compression level in the accuracy of the 3D model and quality of the orthophoto in UAV photogrammetry," *Journal of Imaging*, vol. 6, no. 5, 2020.
- [13] S. Schmid, J. Krabusch, T. Schromm et al., "A new approach for automated measuring of the melt pool geometry in laser-powder bed fusion," *Progress in Additive Manufacturing*, vol. 6, no. 2, pp. 269–279, 2021.
- [14] S. Hegde and S. Gangisetty, "PIG-Net: inception based deep learning architecture for 3D point cloud segmentation," *Computers & Graphics*, vol. 95, pp. 13–22, 2021.
- [15] T. Vrzal, M. Malečková, and J. Olšovská, "DeepRel: deep learning-based gas chromatographic retention index predictor," *Analytica Chimica Acta*, vol. 1147, pp. 64–71, 2021.
- [16] J. Pokhrel and J. Seo, "Statistical model for fragility estimates of offshore wind turbines subjected to aero-hydro dynamic loads," *Renewable Energy*, vol. 163, pp. 1495–1507, 2021.
- [17] K. C. Jung and S. H. Chang, "Advanced deep learning model-based impact characterization method for composite laminates," *Composites Science and Technology*, vol. 207, article 108713, 2021.
- [18] A. Rahman, P. Deshpande, M. S. Radue et al., "A machine learning framework for predicting the shear strength of carbon nanotube-polymer interfaces based on molecular dynamics simulation data," *Composites Science and Technology*, vol. 207, article 108627, 2021.
- [19] W. Zhu, "ANC: attention network for COVID-19 explainable diagnosis based on convolutional block attention module," *Computer Modeling in Engineering & Sciences*, vol. 127, no. 3, pp. 1037–1058, 2021.
- [20] H. Akbari, M. T. Sadiq, and A. U. Rehman, "Classification of normal and depressed EEG signals based on centered correntropy of rhythms in empirical wavelet transform domain," *Health Information Science and Systems*, vol. 9, no. 1, p. 9, 2021.
- [21] M. Rajapandy and A. Anbarasu, "An improved unsupervised learning approach for potential human microRNA-disease association inference using cluster knowledge," *Network Modeling and Analysis in Health Informatics and Bioinformatics*, vol. 10, no. 1, 2021.
- [22] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: visual explanations from deep networks via gradient-based localization," *International Journal of Computer Vision*, vol. 128, no. 2, pp. 336–359, 2020.
- [23] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2921–2929, Las Vegas, NV, USA, 2016.
- [24] A. Alahmadi, A. Davies, J. Royle et al., "An explainable algorithm for detecting drug-induced QT-prolongation at risk of torsades de pointes (TdP) regardless of heart rate and T-wave morphology," *Computers in Biology and Medicine*, vol. 131, article 104281, 2021.
- [25] C. E. Coipan, T. J. Dallman, D. Brown et al., "Concordance of SNP- and allele-based typing workflows in the context of a large-scale international Salmonella Enteritidis outbreak investigation," *Microbial Genomics*, vol. 6, no. 3, article 000318, 2020.
- [26] I. Ali and P. A. Kalra, "A validation study of the 4-variable and 8-variable kidney failure risk equation in transplant recipients in the United Kingdom," *Bmc Nephrology*, vol. 22, no. 1, p. 57, 2021.



- [27] A. Wubalem, "Landslide susceptibility mapping using statistical methods in Uatzau catchment area, northwestern Ethiopia," *Geoenvironmental Disasters*, vol. 8, no. 1, pp. 1–21, 2021.
- [28] Y. Zhang, S. Wang, K. Xia, Y. Jiang, and P. Qian, "Alzheimer's disease multiclass diagnosis via multimodal neuroimaging embedding feature selection and fusion," *Information Fusion*, vol. 66, pp. 170–183, 2021.
- [29] Y. Zhang, F. L. Chung, and S. Wang, "Clustering by transmission learning from data density to label manifold with statistical diffusion," *Knowledge-Based Systems*, vol. 193, article 105330, 2020.
- [30] Y. Zhang, F. L. Chung, and S. Wang, "Fast exemplar-based clustering by gravity enrichment between data objects," *IEEE Transactions on Systems Man Cybernetics-Systems*, vol. 50, no. 8, pp. 2996–3009, 2020.
- [31] H. M. Salama, M. Z. A. el Mageed, G. I. M. Salama, and K. M. Badran, "CSMCSM," *International Journal of Information Security and Privacy*, vol. 15, no. 1, pp. 44–64, 2021.

## Research Article

# Towards Effective Classification of aMCI Based on Resting-State Multiscale Brain Features and Machine Learning Approaches

Chunting Cai <sup>1,2</sup>, Jiqiang Yan,<sup>1</sup> Yu Zhou,<sup>1</sup> Wuyang Zheng <sup>3</sup>, Chenhui Yang <sup>1</sup>,  
Zhemín Zhang <sup>1</sup>, Bokui Chen,<sup>4,5</sup> and Dan Hong<sup>1</sup>

<sup>1</sup>School of Informatics, Xiamen University, Xiamen 361000, China

<sup>2</sup>National Institute for Data Science in Health and Medicine, Xiamen University, Xiamen 361000, China

<sup>3</sup>The First Affiliated Hospital of Xiamen University, Xiamen 361000, China

<sup>4</sup>Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen 518000, China

<sup>5</sup>Artificial Intelligence Research Center, Peng Cheng Laboratory, Shenzhen 518000, China

Correspondence should be addressed to Wuyang Zheng; bamboo789@xmu.edu.cn, Chenhui Yang; ych987@126.com, and Zhemín Zhang; zhangzhemin@xmu.edu.cn

Received 31 March 2021; Revised 26 April 2021; Accepted 30 May 2021; Published 28 June 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Chunting Cai et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Smart healthcare has undergone new opportunities and challenges with the arrival of the Industry 4.0 era. The intelligent imaging diagnosis system is a staple part of smart healthcare, helping doctors make clinical decisions. Nevertheless, intelligent diagnosis analysis is still confronted with the issue that it is challenging to extract effective features from the limited and high-dimensional data, particularly in resting-state data of amnesic mild cognitive impairment (aMCI). Furthermore, the intelligent imaging diagnosis system for aMCI is conducive to make timely predicting groups that may convert to Alzheimer's disease (AD). To improve the system's detection performance and reduce its data redundancy, we first develop an adaptive structure feature generation strategy (ASFGS) based on the Laplacian matrix and sparse autoencoder to obtain the structural features of brain functional network (BFN). Concurrently, we present a multiscale local feature detection strategy (MLFDS) to overcome the low utilization of local features of BFN. And finally, multiscale features, including structural features and multiscale local features, are fused by concatenation method to further improve the detection performance of aMCI system. Support vector machine based on radial basis function (RBF-SVM) for small data learning is adopted to evaluate the effectiveness of the proposed features. Besides, we employ leave-one-out cross-validation strategy to avoid the overfitting problem of classifier training process. The experiment results elucidate that the accuracy (ACC) and the area under the curve (AUC) in this work provide 86.57% and 86.36%, respectively, which outperforms the traditional methods and offers new insights for accuracy requirements of the aMCI system.

## 1. Introduction

Industry 4.0, represented by improvement of the intelligent level of the manufacturing industry, is profoundly converting all walks of life. Smart healthcare that adopts various Industry 4.0 concepts is an era full of opportunities and challenges [1, 2]. As a whole, smart healthcare consists of three parts, including the smart hospital system, family health system, and regional health system [3]. Among them, the core work of a smart hospital system is to collect, store, and process

patients' health status and medical information [4]. Furthermore, imaging diagnosis using medical information and intelligent algorithms can be employed to uncover the risk of disease, timely remind doctors, and assist doctors in making clinical decisions, which is an essential ingredient of smart healthcare [5].

Nowadays, brain-related diseases are considered as one of the most severe problems in the healthcare system. Alzheimer's disease (AD), which frequently occurs in the elderly population, is a disease accompanied by cognitive decline

and noncognitive mental symptoms [6, 7]. Unfortunately, there are no specific drugs or treatment protocols when it comes to AD disease [8]. Moreover, amnesic mild cognitive impairment (aMCI), conceptualized as an episodic memory disorder, is most likely to develop AD [9]. In practice, numerous studies have shown that resting-state functional magnetic resonance imaging (rs-fMRI), characterized by the indirect reflection of neural activity in the brain, is a noninvasive imaging technology that has been widely employed in the classification of brain-related diseases [10, 11]. Accordingly, research on an efficient and reliable system for detecting aMCI is conducive to screening and detecting individuals at high risk for developing AD. It is worth noticing that one of the cores of smart healthcare development is the high demand for data, while aMCI data based on rs-fMRI is confronted with enormous challenges due to its limited data and high dimensions [12].

In this work, the correlation value between the time series of the standard brain regions is calculated using the Pearson correlation coefficient, thus constructing the brain functional network (BFN) that reflects the interaction between the nodes. Significantly, the existing methods only use the local features of the BFN as the input of classifier while ignoring its structural features. To address this issue, we develop an adaptive structure feature generation strategy (ASFGS) based on the Laplacian matrix and sparse autoencoder to improve the classification performance and reduce data redundancy of the system. Concurrently, we present a multiscale local feature detection strategy (MLFDS) to overcome the low utilization of local features of BFN. Afterwards, multiscale features, including structural features and multiscale local features, are fused to further improve classification accuracy of aMCI. It is worth mentioning that support vector machine based on radial basis function (RBF-SVM) for small data learning is utilized to evaluate the performance of the proposed algorithm. In the following, we employ the leave-one-out cross-validation strategy to avoid the overfitting problem of classifier.

Accordingly, the innovativeness of our work is that we first present an ASFGS algorithm to obtain the structural features of BFN, improve the detection accuracy, and reduce data redundancy of the system. Then, we develop an MLFDS algorithm to excavate the local features of BFN at multiple scales. Finally, multiscale features of BFN obtained from the ASFGS algorithm and MLFDS algorithm are concatenated to further improve classification accuracy of aMCI. The results elucidate that the accuracy (ACC) and the area under the curve (AUC) in this work provide about 86.57% and 86.36%, respectively, which outperforms the state-of-the-art methods. It can be inferred that our work dramatically improves the detection performance of aMCI system, providing a new perspective for the construction of intelligent imaging diagnosis system in smart healthcare.

The rest of the work is structured as follows: In Section 2, we review the related works on the feature extraction and classification of aMCI based on rs-fMRI data. In Section 3, we present materials and methods of aMCI detection system. Experiment results and analysis is conducted in Section 4. We conclude the whole work in Section 5.

## 2. Related Works

The recent development and combination of machine learning, statistical algorithm, and neuroimaging technology offer a new perspective for designing an intelligent imaging diagnosis system, which is a crucial procedure toward smart health. The design of an intelligent imaging diagnosis system mainly includes several parts, including the data generation module, data preprocessing module, feature learning module, classifier training module, and feedback module [13]. With the development of intelligent imaging diagnosis technologies, brain network constructed using Pearson correlation coefficient based on rs-fMRI can be employed to estimate the mechanism of information processing and mental expression in the brain, which further proves that it is effective in assisting diagnosis [14, 15]. Nevertheless, owing to the limited and high-dimensional data, little is known about whether to develop the multiscale features of BFN to improve the classification performance of aMCI system. Accordingly, the structural features and multiscale local features that we have developed are the main innovation in this work. This helps us to timely intervene and treat potential individuals associated with brain-related disease.

Numerous works about BFN research have focused on using rs-fMRI to excavate effective features of aMCI. For example, the altered patterns of rich club generated from the BFN have been reported in [16], which indicates that the altered patterns in overlapping nodes can be utilized as the potential features in the aMCI classification process. Moreover, the changes in the architecture of BFN have been reported compared to the healthy control (HC), which is conducive to understanding the mechanism of aMCI and searching for biomarkers [17]. Through the two-sample  $t$ -test, several neuroimaging biomarkers are also identified in the aMCI group, providing a novel aspect for designing interventions before the onset of disease [18]. The preceding studies, restricted by the number of data, are conducted using statistical testing methods to extract the effective features and explore neural mechanisms of aMCI. Furthermore, it ignores the development of the intelligent imaging diagnosis system based on machine learning technology, which is designed to help doctors improve the detection efficiency of aMCI.

With the rapid development of machine learning technology, feature extraction and classification algorithms related to disease have become a hot spot. However, due to the limited number of aMCI data, feature selection is first conducted to reduce redundant information and then use them as the input to classifier to improve classification performance. It provides about the ACC of 69.00% when the significant regional signals resulting from brain pathway activities are employed as the input of support vector machine (SVM) classifier, providing new opportunities for comprehending the disrupted patterns caused by disease [19]. Similarly, the significant features of BFN using the two-sample  $t$ -test are employed to evaluate the classification performance of aMCI in [20], and the result provides about the ACC of 79.10% in SVM classifier. Moreover, it provides about the ACC of 75.35% when the altered signals of low-frequency fluctuation are acted as the input of SVM classifier

[21]. The existing literature extracts the local features through statistical methods, ignoring the multiscale local features and the structural features between nodes of BFN.

In present work, we present an ASFGS algorithm using the Laplacian matrix and sparse autoencoder to obtain the structural features of BFN. Concurrently, we develop an MLFDS algorithm to overcome the low utilization of local features of BFN. In the end, all the features generated above are concatenated to improve the classification performance of aMCI system.

### 3. Materials and Methods

**3.1. Overview of the aMCI Detection System.** The critical point of our work is to design the reliable detection system of aMCI from commonly redundant information of rs-fMRI data, as is shown in Figure 1. To achieve this objective, the proposed aMCI detection system consists of multiple components. At first, the BFN is constructed using the Pearson correlation coefficient. Then, the obtained BFN is utilized as the input of the ASFGS algorithm and MLFDS algorithm we propose to extract features at multiple scales. Furthermore, to evaluate the validity of the proposed algorithm, the RBF-SVM classifier is employed in this project. Ultimately, we send abnormal brain regions and classification results to the doctor in result feedback component.

**3.2. Data Preprocessing.** We utilize the public dataset downloaded from the second phase of Alzheimer's Disease Neuroimaging Initiative (ADNI-2) to validate effectiveness of the proposed aMCI detection system (downloaded from <http://adni.loni.usc.edu/>). The data are preprocessed using a widely adopted Resting-State fMRI Data Analysis Toolkit plus (RESTplus) toolbox [22]. All subjects are required to lie flat, not think, and not move their heads during the scan. It deserves to be noticed that several data do not conform to the basic requirements of the work, such as undue head movement, the reliability of the data, and poor quality of image registration. Specific requirements can be inquired from [18, 23]. Hence, we end up with data on 33 individuals with aMCI and 34 healthy controls (HCs) in this work.

**3.3. BFN Construction.** To better extract the time signals of the corresponding brain regions, anatomical automatic labeling (AAL) template is employed to segment the whole brain into 90 regions of interest (ROI) [24]. Following this, a Pearson correlation coefficient matrix for each data, denoted as  $P_{ij}$ , is calculated using the given time series of standard brain regions. Hence, we obtain a BFN of  $90 \times 90$  dimensions for each data, and the generated Pearson correlation coefficient matrix can be calculated by

$$P_{ij} = \frac{\text{cov}(x_i, x_j)}{\sigma_{x_i} \sigma_{x_j}}, \quad (1)$$

where  $x_i$  and  $x_j$  represent the extracted ROI signals of the brain corresponding to the  $i$  region and  $j$  region. The  $i$  and  $j$  in  $P_{ij}$  represent the brain region positions corresponding

to the AAL template, and each value in the  $P_{ij}$  represents the value of the Pearson correlation coefficient. Moreover,  $\text{cov}(x_i, x_j)$  represents the covariance of variable  $x_i$  and variable  $x_j$ . Likewise,  $\sigma_{x_i}$  and  $\sigma_{x_j}$  represent the standard deviation of variable  $x_i$  and variable  $x_j$ .

**3.4. Adaptive Structure Feature Generation Strategy (ASFGS).** We develop an ASFGS algorithm for extracting the structural features of BFN, which is aimed at improving the detection performance of aMCI system. ASFGS algorithm is mainly composed of two components, including the rough feature extraction module and accurate feature extraction module, as shown in Figure 2.

Considering the substantial contribution of Laplacian Eigenmaps (LE) to maintain and reflect the local relationship between data to some extent, we present the rough feature extraction module based on its conception to extract the structural features of BFN. A brief description of the LE algorithm is as follows [25]:

*Step 1.* Given a set of data  $S = \{s_1, s_2, \dots, s_n\}$ . Where  $s_i$  represents the input data  $i$ , the dimension of each data is denoted by  $q$ , and  $n$  represents the number of all data. In general, Gaussian kernel function is employed to construct an undirected graph matrix between data to depict the adjacency relationship between the data, which can be denoted as  $M_{ij}$ .

$$M_{ij} = e^{-\|s_i - s_j\|^2 / t}. \quad (2)$$

Notably, the degree matrix represents the sum of each column or row in  $M_{ij}$ , which can be calculated as  $D$  through (3). Then, the Laplacian matrix, denoted as  $L$ , can be obtained by (4).

$$D = \sum_j M_{ij}, \quad (3)$$

$$L = D - M. \quad (4)$$

*Step 2.* Since the Laplace matrix is a positive semidefinite matrix, it can be further expressed as:

$$\begin{aligned} y^T L y &= y^T D y - y^T M y = \sum_{i=1}^n d_i y_i^2 - \sum_{i,j=1}^n y_i y_j m_{ij} \\ &= \frac{1}{2} \left( \sum_{i=1}^n d_i y_i^2 - 2 \sum_{i,j=1}^n y_i y_j m_{ij} + \sum_{j=1}^n d_j y_j^2 \right) \\ &= \frac{1}{2} \sum_{i,j=1}^n m_{ij} (y_i - y_j)^2 \geq 0. \end{aligned} \quad (5)$$

To maintain the adjacency relation between two data, it can be converted to the minimization issue. That is, if  $m_{ij}$  is larger, the higher the similarity between  $y_i$  and  $y_j$ . Thus, it is now reduced to

$$y_{\text{opt}} = \arg \min_{y^T D y = 1} y^T L y. \quad (6)$$

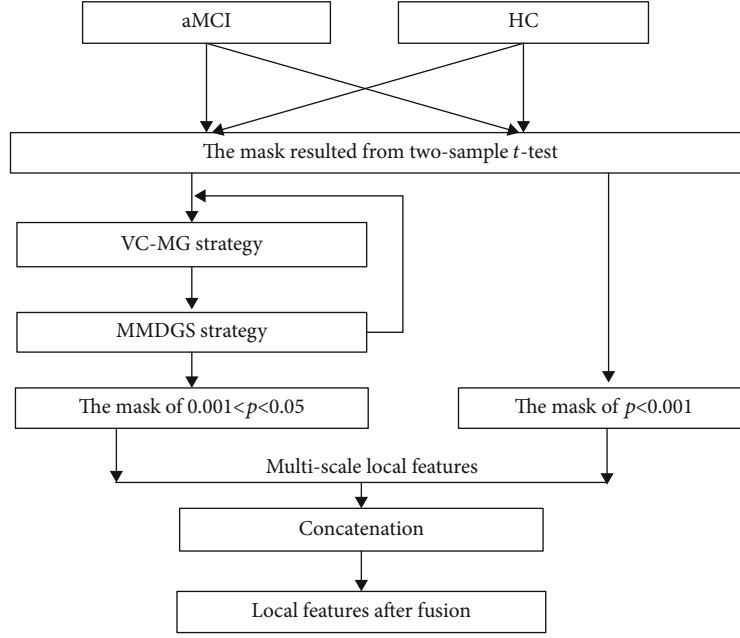


FIGURE 1: Overall framework of aMCI detection system.

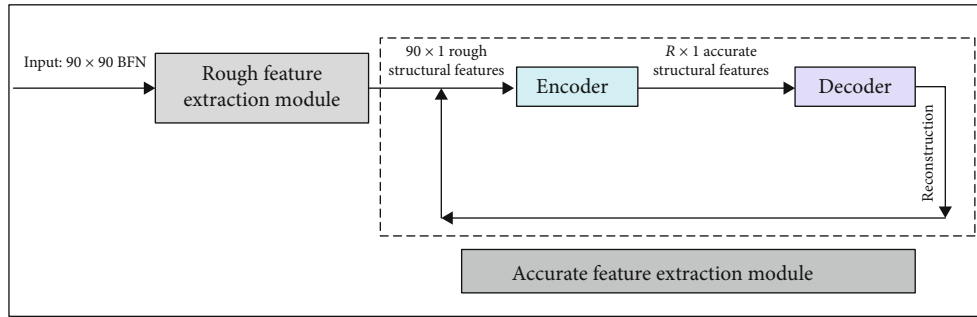


FIGURE 2: ASFGS algorithm for extracting structural features of BFN.

Step 3. Ultimately, by employing the Lagrange multiplier method, it can be approximately converted into

$$Ly = \lambda Dy. \quad (7)$$

Nevertheless, the objective of the LE algorithm is to reduce the dimension of data features, while the retained dimension is determined by the number of minimum nonzero eigenvalues of the matrix, which does not conform to the requirements of the structural feature extraction of BFN in this paper. Fortunately, LE algorithm plays a significant role in maintaining the relation between sample points after dimensionality reduction. Therefore, we present a rough feature extraction module that modifies the LE algorithm to put its proper focus on structural feature extraction of BFN.

First of all, the minimization problem is constructed using the Laplacian matrix [25], as is shown in

$$\min y^T Ly = \frac{1}{2} \sum_{i,j=1}^B m_{ij} (y_i - y_j)^2 \quad (8)$$

$$\text{s.t. } y^T Dy = 1,$$

where  $B$  represents the number of brain nodes. To better evaluate the interrelation between brain region signals in BFN, we use the Pearson correlation coefficient matrix obtained by (1) to replace the Gaussian kernel function. It is worth noting that at this point  $D$  represents the level of importance of brain nodes.

Considering the limitation of the small data set, we reduce the dimension of BFN from  $90 \times 90$  to  $90 \times 1$ , while maintaining the inherent correlation between brain nodes of the BFN. Let  $(\lambda_f, y_f)$  denotes the solution to (7). Hence, the rough feature extraction module can be induced as follows:

$$\arg \min_{y^T Dy=1} y^T Ly = \lambda_f Dy, y = y_f \longrightarrow \min \lambda_f y_f^T Dy, f \in (1, 2, \dots, h), \quad (9)$$

where  $h$  represents the number of nonzero eigenvalues of the  $90 \times 90$  matrix of BFN. To summarize, the optimal result of (8) can be obtained by (9). The novel aspect of our work is that we present the rough feature extraction module to extract the structural features of BFN under the condition of the small data set. While the generated vector still contains redundant

information, such as the randomness of the values of unrelated brain regions. Based on this, the accurate feature extraction module is presented using the sparse autoencoder to further extract the structural features of BFN in this work. Sparse autoencoder, which constrains hidden layer neurons to some extent and continuously regulates parameters through the errors between the output and input of the model, is an unsupervised machine learning algorithm [26]. Besides, the sparse autoencoder can represent linear or nonlinear transformations using different transfer functions in the coding process. The advantages of using the sparse autoencoder are as follows: Firstly, it can reduce the dimensionality of the original data while maintaining the nonlinear structural features of the data. Secondly, the feature dimension extracted using the sparse autoencoder is not fixed, and the structure of the input data can be obtained by regulating parameters to minimize model error. In the end, various brain regions are inhibited or activated, respectively, in the resting state, while the neurons in the hidden layer of the sparse autoencoder are sparse, indicating that the sparse autoencoder is desirable for the actual situation in which the human brain works. Accordingly, we present an accurate feature extraction module to further extract the structural features and reduce redundant information.

The accurate feature extraction module consists of a hidden layer, and the transfer functions of the encoder and decoder are nonlinear. First, the cost function of the sparse autoencoder using sparse constraint in the hidden layer is given as follows [27, 28]:

$$\begin{aligned}
J(W, b) = & \frac{1}{\mu} \sum_{i=1}^{\mu} \left( \frac{1}{2} \left\| h_{W,b}(x^{(i)}) - y^{(i)} \right\|^2 \right) \\
& + \frac{\lambda}{2} \sum_{l=1}^{L-1} \sum_{i=1}^{s_l} \sum_{j=1}^{s_{l+1}} \left( \theta_{ji}^{(l)} \right)^2 \\
& + \beta \sum_{j=1}^{s_2} \text{KL}(\rho \parallel \hat{\rho}_j),
\end{aligned} \quad (10)$$

where  $\mu$  represents the number of training data,  $x^{(i)}$  represents the input data,  $y^{(i)}$  represents label of data,  $\lambda$  represents the weight decay parameter,  $\beta$  represents the tuning parameter,  $s^2$  represents the number of neurons in the hidden layer,  $\text{KL}(\bullet)$  represents relative entropy,  $W$  represents the weight coefficient of encoder,  $L$  represents the number of layers, and  $b$  represents the bias coefficient, respectively.

Here, let  $s_2$  denotes the average node degree of all data to represent the average activity level of brain nodes. That is, we substitute the number of active brain regions for the number of neurons in the hidden layer, where  $r$  represents the number of brain nodes.

$$s_2 = \frac{1}{n} \sum_{g=1}^n \frac{1}{r^2} \left( \sum_{i=1}^r \sum_{j=1}^r M_{ij} \right). \quad (11)$$

In order to optimize the error between the output and input of the sparse autoencoder, the back propagation algorithm is employed to update the model parameters. Owing

to the limited number of data, we further employ the two-sample  $t$ -test to obtain the structural features with significant differences ( $p < 0.05$ ) in the  $R \times 1$  generated vector of BFN, where  $R$  represents the number of neurons in hidden layer.

**3.5. Multiscale Local Feature Detection Strategy (MLFDS).** We develop an MLFDS algorithm to overcome the low utilization of local features of BFN. Our algorithm is proposed based on maximizing the mean difference between classes and minimizing the intraclass variance. The framework of the MLFDS algorithm shown in Figure 3 is mainly composed of two parts, including the mask generation based on variable coefficient (VC-MG) and minimal mean difference generation strategy (MMDGS).

At present, the two-sample  $t$ -test is employed to obtain the nodes of BFN with significant difference ( $p < 0.05$ ) in mean value between classes, which is conducive to removing redundant information. However, before the two-sample  $t$ -test is performed, the results of homogeneity analysis can not determine whether the variance of each type of data is large or small. Besides, the latest method only uses the mask of  $p < 0.001$  for extracting local features of BFN, ignoring the effective utilization of features, such as the features under the mask of  $0.001 < p < 0.05$  [20, 21]. Consequently, the MLFDS algorithm is presented to address this issue.

*Step 1.* We first calculate the variable coefficient of BFN in two groups to generate the mask, respectively. The value of corresponding position is 0 if variable coefficient in the mask is greater than the mean; otherwise, it is 1, where 1 means that variation coefficient is lower than the average value of variation coefficient. It is remarkable that we employ the median of variation coefficient as the average value to avoid the influence of extreme values. Then, we intersect the generation masks of the two kinds of data generated by the above operation, and the obtained mask is denoted as  $\text{Mask}_i$ , where the  $i$  represents the number of executions of Step 1 (VC-MG).

*Step 2.* We calculate the average value of  $p$  at the corresponding positions according to the  $\text{Mask}_i$ . Concurrently, we perform the subtraction operation on the variation coefficient matrix between classes according to the  $\text{Mask}_i$  and then calculate its average value  $E_i$  (MMDGS).

*Step 3.* Perform Step 1 and Step 2  $n$  times.

$$E_i = \frac{\text{sum}(|A_i - B_i|)}{l_i}, \quad (12)$$

where  $A_i$  represents the variable coefficient matrix of the aMCI group,  $B_i$  represents the variable coefficient matrix of the HC group,  $l_i$  represents the number of 1 in  $\text{Mask}_i$ ,  $\eta$  is constrained to (13), and  $N$  is equal to 8100.

$$2^n \leq N. \quad (13)$$

The principle of MLFDS algorithm is to select the position with lower variance under the premise of the obvious difference in mean value between groups. Therefore, the  $i$

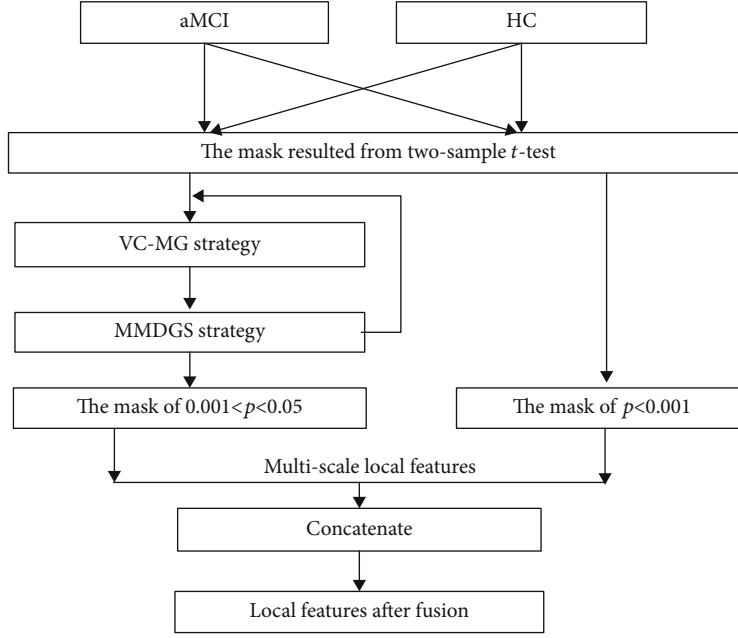


FIGURE 3: The MLFDS algorithm to extract the multiscale local features.

corresponding to the lowest value of  $E_i$  is employed as the threshold value to extract the local features of BFN under the mask of  $0.001 < p < 0.05$ . And finally, the multiscale local features resulted from the mask of  $0.001 < p < 0.05$  and mask of  $p < 0.001$  are concatenated in this work to improve the detection performance.

**3.6. Multiscale Feature Fusion.** In order to improve the detection performance of aMCI system, we concatenate multiscale features resulted from the ASFGS algorithm and MLFDS algorithm in this work. That is, supposing the dimensions of two groups of features are  $d_1$  and  $d_2$ , respectively, then the dimension of concatenated feature is equal to  $d_1 + d_2$ .

**3.7. Classification Using RBF-SVM Classifier.** In view of the finite data, the adoption of suitable classifier is essential to estimate the validation of features obtained from the proposed algorithms. Fortunately, numerous works on mild cognitive impairment (MCI) classification have shown that RBF-SVM classifier has superior detection performance [19–21]. The following is a brief introduction to the RBF-SVM classifier:

The essence of SVM algorithm is to work around the optimization problem of the objective function [29].

$$\min_{w,b} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^u V_i \quad (14)$$

$$\text{s.t. } y_i [w^T x_i + b] \geq 1 - \psi_i, \psi_i \geq 0.$$

The objective of Gaussian radial basis function is to obtain the new space, which is more favourable to classification [30].

$$K(x, x_i) = \exp\left(-\frac{\|x - x_i\|^2}{\sigma^2}\right), \quad (15)$$

where  $C$  denotes the penalty coefficient,  $x_i$  denotes the input data,  $y_i$  denotes the label of data,  $w$  denotes the weight coefficient,  $b$  denotes the bias coefficient, and  $V_i$  denotes the relaxation variable, respectively.

To prevent the overfitting issue of classifier training process, we utilize the leave-one-out cross-validation strategy in this work.

**3.8. Evaluation Criteria.** To measure the performance of the classification model, the frequently used metrics for binary classification are ACC, F1-score, AUC, etc. Significantly, false positive (FP), false negative (FN), true negative (TN), and true positive (TP) are defined using the confusion matrix, as shown in Figure 4 [31, 32].

$$\begin{aligned} \text{ACC} &= \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \\ \text{Precision} &= \frac{\text{TP}}{\text{TP} + \text{FP}}, \\ \text{Sensitivity} &= \frac{\text{TP}}{\text{TP} + \text{FN}}, \\ \text{FI-Score} &= \frac{2 \times \text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}}. \end{aligned} \quad (16)$$

## 4. Experiment Results and Analysis

This work explores the detection power of aMCI system using multiscale features of BFN, which are derived from rs-fMRI data, for the automatic identification and classification of aMCI subjects from HCs. In the proposed detection system, we employ structural features using ASFGS algorithm and multiscale local features using MLFDS algorithm to train an RBF-SVM classifier for accurate discrimination of aMCI individuals.

		Predicted value	
		$y = 1$ , aMCI	$y = 0$ , HC
True value	$y = 1$ , aMCI	True positive (TP)	False negative (FN)
	$y = 0$ , HC	False positive (FP)	True negative (TN)

FIGURE 4: The confusion matrix which is utilized to evaluate the binary classification.

**4.1. The Performance Analysis of ASFGS Algorithm.** The rough feature extraction module is first presented to extract the structural features of BFN, which maintains the correlation between the brain nodes after dimensionality reduction from  $90 \times 90$  to  $90 \times 1$ . Considering the generated vector mentioned above still contains redundant information, we develop an accurate feature extraction module to further extract the structural features and achieve the dimensioning reduction from  $90 \times 1$  to  $12 \times 1$ . Next, the two-sample  $t$ -test ( $p < 0.05$ ) is employed for  $12 \times 1$  structural feature generation vector to reduce information redundancy under the condition of limited data. In the end, the structural features with obvious alteration are employed as the input of SVM-RBF classifier. The classification results using ASFGS algorithm are shown in Table 1. The ASFGS algorithm provides about the ACC of 61.20% in RBF-SVM classifier. Also, it provides about the AUC of 62.12% in RBF-SVM classifier. For F1-score, it achieves about 60.61% performance in RBF-SVM classifier. The metric values of classification are all greater than 60.00%, indicating that the structural features extracted by ASFGS algorithm can improve the detection performance of aMCI.

Our goal is to simulate the information processing pattern of human brain to extract structural information of BFN, which further improves the detection performance of the aMCI system. Consequently, mathematical modelling about BFN is implemented, which contains information about the interactions between brain regions [33–35]. We perform the rough feature extraction module to extract the structural features; that is, the information of brain regions with higher correlation will be maintained after dimensionality reduction. Next, the accurate feature extraction module is based on how the brain works in the resting-state, in which some parts of the brain nodes are activated while others are suppressed. Let the number of neurons in the hidden layer of the sparse autoencoder be the average activity level of brain nodes, while the selection of sparsity can make some brain regions in the inhibited state and others in the activated state. Where 12 is derived from (11), which is performed to represent the average activity level of brain nodes. The results show that the sparsity threshold between 0.4 and 0.5 shown in Figure 5 has the minimum reconstruction error, indicating that the number of activated brain regions is about 4 to 6. The activity level of brain nodes (node degree) in the data ranges from 3 to 31, and the sparsely activated brain regions are also within this range, suggesting that the brain working mechanism we simulate is meaningful to some extent.

TABLE 1: The structural feature analysis using ASFGS algorithm.

Algorithm	ACC	F1-score	AUC
ASFGS	61.20%	60.61%	62.12%

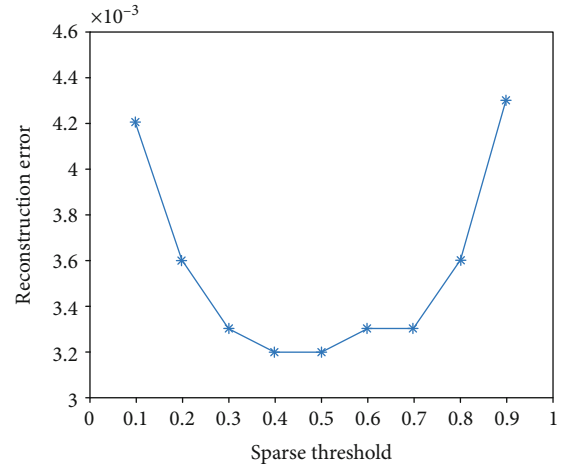


FIGURE 5: Reconstruction error of sparse autoencoder at each threshold.

**4.2. The Performance Analysis of MLFDS Algorithm.** We develop an MLFDS algorithm to excavate the multiscale local features of the BFN. Specifically, we first present the VC-MG strategy to generate the mask, and then, we present the MMDGS strategy based on the mask to extract the multiscale local features.

We extract the local fusion features of BFN using MLFDS<sup>2</sup> algorithm to improve the detection performance of aMCI system. The results elucidate that five pairs of connected brain nodes with obvious alteration are found using MLFDS<sup>1</sup> algorithm, including (21, 72), (45, 46), (11, 61), (73, 76), and (74, 76), as shown in Figure 6. Furthermore, two pairs of connected brain nodes with obvious alteration are found using SLF, including (63, 76) and (58, 64). It is worth noting that several numbers in Figure 6 correspond to specific brain regions in the AAL template, which can be found in [36]. Where the blue ball denotes the brain nodes with obvious alteration, the red lines show the great correlation in two brain nodes, SLF refers to the single local feature method (two-sample  $t$ -test,  $p < 0.001$ ) in [20, 21], MLFDS<sup>1</sup> refers to local features obtained using the VC-MG strategy and MMDGS strategy, and MLFDS<sup>2</sup> refers to features after fusion of multiscale local features.

As shown in Table 2, our findings elucidate that using MLFDS<sup>1</sup> algorithm can achieve about the ACC of 79.10% in RBF-SVM classifier, and the improvement is 2.98% compared to SLF algorithm. Also, the AUC of RBF-SVM classifier is 79.14%, increasing by 2.67% compared to SLF algorithm. For F1-score metric, it provides about 77.42% performance, increasing by 1.66% compared to SLF algorithm. From the classification results, the MLFDS<sup>1</sup> algorithm is more effective than the state-of-the-art algorithm (SLF). This is due to the fact that the MLFDS<sup>1</sup> algorithm follows the



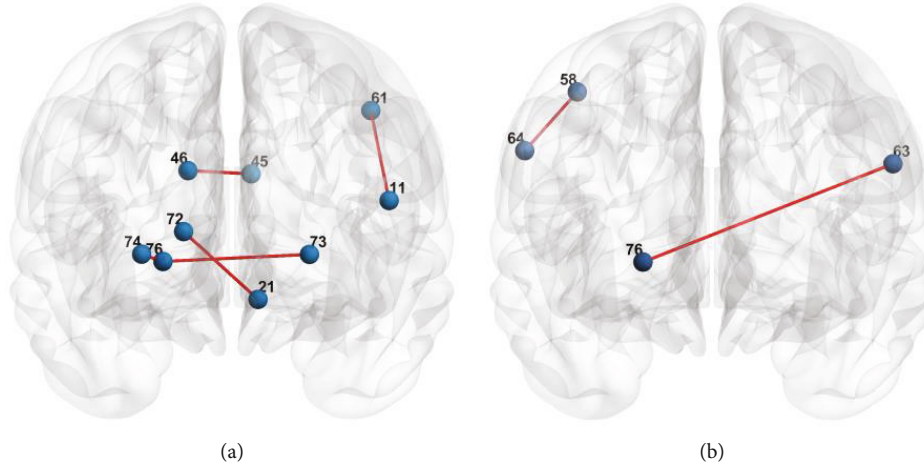


FIGURE 6: The brain nodes involved in multiscale local feature selection: (a) MLFDS<sup>1</sup> algorithm; (b) SLF algorithm.

TABLE 2: The multiscale local feature analysis using MLFDS algorithm.

Algorithm	ACC	F1-score	AUC
SLF	76.12%	75.76%	76.47%
MLFDS <sup>1</sup>	79.10%	77.42%	79.14%
MLFDS <sup>2</sup>	80.60%	80.00%	84.22%

TABLE 3: The fusion analysis of structural features and multiscale local features.

Algorithm	ACC	F1-score	AUC
SLF	76.12%	75.76%	76.47%
ASFSGS-MLFDS	86.57%	85.71%	86.36%

principle of maximizing the mean difference between classes and minimizing the intraclass variance. Therefore, the discriminant features used for classification can be obtained to some extent. We further concatenate the multiscale local features generated from MLFDS<sup>1</sup> algorithm and SLF algorithm. The results elucidate that using MLFDS<sup>2</sup> algorithm provides about the ACC of 80.60% in RBF-SVM classifier, increasing by 4.48% compared to SLF algorithm. Moreover, it achieves about 84.22% in AUC, with an improvement rate of 7.75% compared to SLF algorithm. For F1-score metric, it provides about 80.00% performance in RBF-SVM classifier, increasing by 4.24% compared to SLF algorithm. This indicates that the concatenation of multiscale local features can greatly improve the detection performance of aMCI system.

**4.3. Performance Analysis of Fusion of Structural Features and Multiscale Local Features.** In order to further improve the detection of aMCI system, we mainly concatenate the multiscale features of BFN, referred to as ASFSGS-MLFDS, including structural features and multiscale local features.

As is shown in Table 3, ASFSGS-MLFDS algorithm provides about the ACC of 86.57% in RBF-SVM classifier, increasing by 10.45% compared to SLF algorithm. Besides, it provides about the AUC of 86.36% in RBF-SVM classifier, with an improvement rate of 9.89% compared to SLF algo-

rithm. For F1-score, it provides about 85.71% performance in RBF-SVM classifier, increasing by 9.95% compared to SLF algorithm. This elucidates that multiscale local features and structural features play a complementary role, which significantly improves the detection performance of aMCI system, thus making up for the low feature utilization rate under the condition of limited data.

## 5. Conclusions

In this paper, we develop an aMCI detection system. Firstly, we present the ASFSGS algorithm to extract structural features of BFN. Then, we present the MLFDS algorithm that excavates the multiscale local features of BFN, thus overcoming the low utilization of local features. In the end, multiscale features of BFN, including structural features and multiscale local features, are fused to further improve the detection performance of aMCI system. Our work outperforms the state-of-the-art methods and offers new insights for the accuracy requirement of aMCI system. Accordingly, the ASFSGS algorithm and MLFDS algorithm we present can be employed to detect brain diseases, providing new insights for the intelligent construction of the imaging diagnosis system. The future work contains introducing multimodality data to improve the detection performance of aMCI system.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work has been supported by the Natural Science Foundation of Fujian Province of China (No. 2017J01372), the Foundation of Fujian Educational Committee (No.

JK2015019), the National 135 Key R & D Program Projects (Grant No. 2018YFB1600600), the Tsinghua Overseas Research Cooperation Project (Grant No. HW2020005), and the Science and Technology Innovation Committee of Shenzhen Project (Grant No. JCYJ20190813173401651).

## References

- [1] P. Pace, G. Aloï, R. Gravina, G. Caliciuri, G. Fortino, and A. Liotta, "An edge-based architecture to support efficient applications for healthcare Industry 4.0," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 1, pp. 481–489, 2019.
- [2] H. Qiu, M. Qiu, M. Liu, and G. Memmi, "Secure health data sharing for medical cyber-physical systems for the healthcare 4.0," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 9, pp. 2499–2505, 2020.
- [3] A. Alaiad and L. Zhou, "Patients' adoption of WSN-based smart home healthcare systems: an integrated model of facilitators and barriers," *IEEE Transactions on Professional Communication*, vol. 60, no. 1, pp. 4–23, 2017.
- [4] H. Zhang, J. Li, B. Wen, Y. Xun, and J. Liu, "Connecting intelligent things in smart hospitals using NB-IoT," *IEEE Internet of Things Journal*, vol. 5, no. 3, pp. 1550–1560, 2018.
- [5] S. U. Amin, M. S. Hossain, G. Muhammad, M. Alhussain, and M. A. Rahman, "Cognitive smart healthcare for pathology detection and monitoring," *IEEE Access*, vol. 7, pp. 10745–10753, 2019.
- [6] Y. Zhao, B. Ma, P. Jiang, D. Zeng, X. Wang, and S. Li, "Prediction of Alzheimer's disease progression with multi-information generative adversarial network," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 3, pp. 711–719, 2021.
- [7] T. Tong, Q. Gao, R. Guerrero et al., "A novel grading biomarker for the prediction of conversion from mild cognitive impairment to Alzheimer's disease," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 1, pp. 155–165, 2017.
- [8] P. Sundaram, M. Luessi, M. Bianciardi, S. Stufflebeam, M. Hämäläinen, and V. Solo, "Individual resting-state brain networks enabled by massive multivariate conditional mutual information," *IEEE Transactions on Medical Imaging*, vol. 39, no. 6, pp. 1957–1966, 2020.
- [9] Y. Li, J. Liu, Z. Tang, and B. Lei, "Deep spatial-temporal feature fusion from adaptive dynamic functional connectivity for MCI identification," *IEEE Transactions on Medical Imaging*, vol. 39, no. 9, pp. 2818–2830, 2020.
- [10] S. Liu, S. Liu, W. Cai et al., "Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 4, pp. 1132–1140, 2015.
- [11] R. Li, G. Rui, C. Zhao, C. Wang, F. Fang, and Y. Zhang, "Functional network alterations in patients with amnesic mild cognitive impairment characterized using functional near-infrared spectroscopy," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 1, pp. 123–132, 2020.
- [12] D. Wen, P. Li, Y. Zhou et al., "Feature classification method of resting-state EEG signals from amnesic mild cognitive impairment with type 2 diabetes mellitus based on multi-view convolutional neural network," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 8, pp. 1702–1709, 2020.
- [13] C. Lin, Y. Teng, J. J. M. Tan, and L. Hsu, "Local diagnosis algorithms for multiprocessor systems under the comparison diagnosis model," *IEEE Transactions on Reliability*, vol. 62, no. 4, pp. 800–810, 2013.
- [14] Z. Wang, Y. Zheng, D. C. Zhu, A. C. Bozoki, and T. Li, "Classification of Alzheimer's disease, mild cognitive impairment and normal control subjects using resting-state fMRI based network connectivity analysis," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 6, pp. 1–9, 2018.
- [15] Q. Zhou, M. Goryawala, M. Cabrerizo et al., "An optimal decisional space for the classification of Alzheimer's disease and mild cognitive impairment," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 8, pp. 2245–2253, 2014.
- [16] C. Xue, H. Sun, G. Hu et al., "Disrupted patterns of rich-club and diverse-club organizations in subjective cognitive decline and amnesic mild cognitive impairment," *Frontiers in Neuroscience*, vol. 14, 2020.
- [17] M. Filippi, S. Basaia, E. Canu et al., "Changes in functional and structural brain connectome along the Alzheimer's disease continuum," *Molecular Psychiatry*, vol. 25, no. 1, pp. 230–239, 2020.
- [18] C. Cai, C. Huang, C. Yang et al., "Altered patterns of functional connectivity and causal connectivity in salience subnetwork of subjective cognitive decline and amnesic mild cognitive impairment," *Frontiers in Neuroscience*, vol. 14, 2020.
- [19] J. Lee, Y. Kim, Y. Jeong et al., "Inference of brain pathway activities for Alzheimer's disease classification," *BMC Medical Informatics and Decision Making*, vol. 15, Supplement 1, 2015.
- [20] X. Zhang, B. Hu, X. Ma, and L. Xu, "Resting-state whole-brain functional connectivity networks for MCI classification using L2-regularized logistic regression," *IEEE Transactions on NanoBioscience*, vol. 14, no. 2, pp. 237–247, 2015.
- [21] L. Yang, Y. Yan, Y. Wang et al., "Gradual disturbances of the amplitude of low-frequency fluctuations (ALFF) and fractional ALFF in Alzheimer spectrum," *Frontiers in Neuroscience*, vol. 12, p. 975, 2018.
- [22] X. Z. Jia, J. Wang, H. Y. Sun et al., "Restplus: an improved toolkit for resting-state functional magnetic resonance imaging data processing," *Science Bulletin*, vol. 64, pp. 953–954, 2019.
- [23] C. Chunting, H. Chenxi, C. Yang et al., "Altered patterns of phase position connectivity in default mode subnetwork of subjective cognitive decline and amnesic mild cognitive impairment," *Frontiers in Neuroscience*, vol. 14, 2020.
- [24] R. Ju, C. Hu, P. Zhou, and Q. Li, "Early diagnosis of Alzheimer's disease based on resting-state brain networks and deep learning," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 16, no. 1, pp. 244–257, 2019.
- [25] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," *Advances in Neural Information Processing Systems*, vol. 14, no. 6, pp. 585–591, 2001.
- [26] Y. Su, A. Marinoni, J. Li, J. Plaza, and P. Gamba, "Stacked non-negative sparse autoencoders for robust hyperspectral unmixing," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 9, pp. 1427–1431, 2018.
- [27] Z. Shao, L. Wang, Z. Wang, and J. Deng, "Remote sensing image super-resolution using sparse representation and coupled sparse autoencoder," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 8, pp. 2663–2674, 2019.

- [28] F. H. C. Tivive and A. Bouzerdoum, "Clutter removal in through-the-wall radar imaging using sparse autoencoder with low-rank projection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 2, pp. 1118–1129, 2021.
- [29] F. Fang, H. Tian, and Y. Li, "SVM strategy for mitigating low-order harmonics in isolated AC–DC matrix converter," *IEEE Transactions on Power Electronics*, vol. 36, no. 1, pp. 583–596, 2021.
- [30] S. Asaly, L.-A. Gottlieb, and Y. Reuveni, "Using support vector machine (SVM) and ionospheric total electron content (TEC) data for solar flare predictions," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 1469–1481, 2021.
- [31] T.-E. Kam, H. Zhang, Z. Jiao, and D. Shen, "Deep learning of static and dynamic brain functional networks for early MCI detection," *IEEE Transactions on Medical Imaging*, vol. 39, no. 2, pp. 478–487, 2020.
- [32] Y. Zhang, S. Wang, K. Xia, Y. Jiang, and P. Qian, "Alzheimer's disease multiclass diagnosis via multimodal neuroimaging embedding feature selection and fusion," *Information Fusion*, vol. 66, pp. 170–183, 2021.
- [33] Y. Shi, W. Zeng, J. Deng, W. Nie, and Y. Zhang, "The identification of Alzheimer's disease using functional connectivity between activity voxels in resting-state fMRI data," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 8, pp. 1–11, 2020.
- [34] P. Yang, F. Zhou, D. Ni et al., "Fused sparse network learning for longitudinal analysis of mild cognitive impairment," *IEEE Transactions on Cybernetics*, vol. 51, no. 1, pp. 233–246, 2021.
- [35] Y. Zhang, F. L. Chung, and S. Wang, "Clustering by transmission learning from data density to label manifold with statistical diffusion," *Knowledge-Based Systems*, vol. 193, article 105330, 2020.
- [36] R. Salvador, J. Suckling, M. R. Coleman, J. D. Pickard, D. Menon, and E. Bullmore, "Neurophysiological architecture of functional magnetic resonance images of human brain," *Cerebral Cortex*, vol. 15, no. 9, pp. 1332–1342, 2005.

## Research Article

# Multiobjective Optimization regarding Vehicles and Power Grids

Kaiyang Zhong <sup>1</sup>, Ping Wang,<sup>1</sup> Jiaming Pei <sup>2</sup>, Jiuyan Xu <sup>2</sup>, Zonglin Han <sup>3</sup>,  
and Jiawen Xu <sup>4</sup>

<sup>1</sup>School of Economic Information Engineering, Southwestern University of Finance and Economics, Chengdu, Sichuan, China

<sup>2</sup>School of Computer Science and technology, Taizhou University, Jiangsu, China

<sup>3</sup>Zhengzhou Electric Power Vocational and Technical College, Zhenzhou, China

<sup>4</sup>College of Artificial Intelligence, Tianjin University of Science and Technology, Tianjin, China

Correspondence should be addressed to Jiaming Pei; [jiamingpei0262@gmail.com](mailto:jiamingpei0262@gmail.com)

Received 11 April 2021; Accepted 2 June 2021; Published 14 June 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Kaiyang Zhong et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Vehicle to Grid (V2G) refers to the optimal management of the charging and discharging behavior of electric vehicles through reasonable strategies and advanced communication. In the process of interaction, there are three stakeholders: the power grid, operators (charging stations), and EV users. In real life, the impact of peak-valley difference caused a lot of power loss when charging. At the same time, the loss of current is also a loss for power grid companies and EV users. In this paper, we propose a multiobjective optimization method to reduce the current loss and determine the relationship between the parameters and the objective function and constraints. This optimization method uses a genetic algorithm for multiobjective optimization. Through the analysis of the number of vehicles and load curve of AC class I and AC class II electric vehicles before and after optimization in each period, we found that the charging load of electric vehicles played a role of valley filling in the low valley price stage and played a peak-cutting role in a peak price period.

## 1. Introduction

Automobile energy consumption accounts for nearly a quarter of the world's total energy consumption. With the improvement of the economic level of developing countries [1], the number of automobiles is increasing sharply, and the energy and environmental problems caused are becoming more serious [2]. Governments around the world are actively promoting electric vehicles. In addition to a series of economic subsidy policies regarding production and sales, many countries, including China, have also proposed the agenda the formulation of a timetable to ban the sale of fuel-fired vehicles. Electric vehicles (EVs) are considered a solution to reduce air pollution and greenhouse gas emissions. As a result, their market share has increased exponentially in recent years [3]. The energy of electric vehicles mainly comes from the power grid, which brings challenges to the distribution network, such as the increase of the peak-valley load difference, voltage drop, and loss increase. However, electric vehicles have broad appli-

cation prospects in peak shaving, valley filling, power system auxiliary services, and the coordinated absorption of new energy as a mobile energy storage [4]. The main tasks are allocating, setting up charging stations, and regulating the timing of the photovoltaic power generation and power supply, which can minimize user costs and the loss of the power grid.

Many scholars have proposed solutions to the problem of electric vehicles from different angles. Miao et al. [5] introduced and compared the key components of lithium-ion batteries and described the related battery management system and the methods for improving the overall battery efficiency, capacity, and life. Researchers [6, 7] studied the energy management strategies and charging technology for pure electric vehicles, discussed the main challenges and corresponding solutions faced by pure electric vehicles, and introduced the latest developments in pure electric vehicles.

However, research and political interest in public charging has increasingly focused on fast charging options and high electricity prices, but few estimates of future demand.

Gnann et al. [8] contributed to the research field related to policy by determining the number of fast charging stations required for each electric vehicle in the queuing model and the potential additional electricity saved per kilowatt hour. Furthermore, Pei used the PSO of intelligent computing to solve the charging and discharging theory of electric vehicles earlier [9], which provided inspiration.

In this paper, our main contribution is to propose a multiobjective optimization method to reduce the current loss of electric vehicles during the charging process, which is of benefit to both electric power companies and electric vehicle owners. The work we have done is as follows:

- (1) Select the Monday charging record to arrange. As the charging connection time is the physical connection time between the electric vehicle and the charging pile, the charging power is divided by the charging connection time, and the maximum power in a week is regarded as the minimum power of the car, which is sorted by power from small to large
- (2) With the goal of minimizing user charging costs, the optimal allocation plan for 5000 vehicles in a 24-hour period is obtained. In addition to the power limit of the line, the vehicle's own charging and discharging power limit, the battery SOC limit, etc. are constraints, and the final power of each vehicle is 50 kWh

## 2. Problem Description

A new energy charging station provides charging service for electric vehicles by utilizing grid energy and new energy in the station and can send the new energy in the station back to the grid. Figure 1 shows the system structure of a new energy charging station. There are eight charging piles and photovoltaic power generation devices in the station. The charging power of each charging pile can be adjusted in the range of 50 kW [10]. Some symbol descriptions are shown in Table 1.

When electric vehicles are parked in residential areas, this often exchanges electricity directly with the distribution network through charging piles (i.e., charging or discharging), which may lead to a significant change in the distribution of power flow.

Figure 2 shows the distribution network topology of a residential area.

Assuming that node 8 is connected to 100 electric vehicles, 10% of them are charged by AC level 1 and 90% are charged by AC level 2. The discharge power is not more than 5 kW. The battery capacity of the electric vehicle is 50 kWh, and the state of charge (SOC) of the battery is no less than 0.2 [11].

We must provide the optimal charging and discharging scheme to minimize the active power loss for the power grid and the total cost of electric vehicle owners by establishing an optimization model under the circumstance where the optimal period is 24 hours a day.

## 3. Methods and Experiments

*3.1. Classification Model of Electric Vehicle Charging Power Level.* Given the charging power meter of 100 electric vehicles

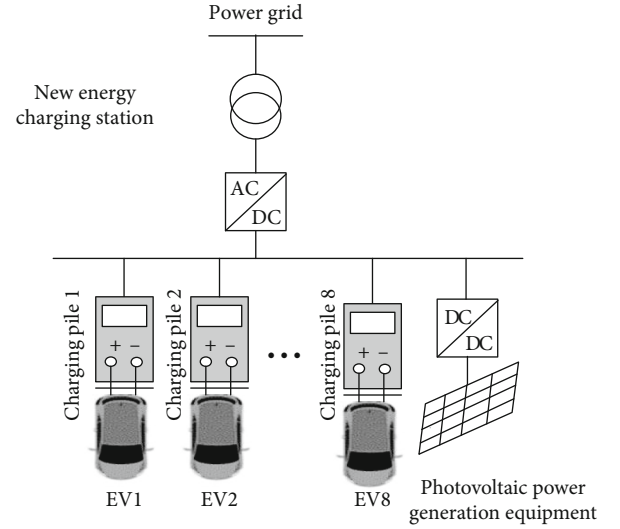


FIGURE 1: New energy charging station structure.

in a city, we can calculate the minimum power of electric vehicles in a day. The function is as follows:

$$P_{\min} = \frac{w}{t_{\max}}. \quad (1)$$

We used an Excel function to process the data and used the automatic sorting function in Excel to ascend the minimum charging power and classify the top 10 vehicles into AC class 1 (1.4-1.9 kW). We, thus, obtained Table 2.

*3.2. Drawing the 24-Hour Charging Load Curve.* To more accurately calculate the actual charging time of electric vehicles according to the data of Table 2, the average value of the power interval is

$$P_i = \frac{P_{i1} + P_{i2}}{2}. \quad (2)$$

After substituting the data in Table 1, we can obtain the charging power of AC 1 at  $P_1 = 1.65$  kW, the power of AC 2 is  $p_2 = 33.3$  kW, and DC power is  $p_3 = 70$  kW.

According to the energy formula, the actual charging time for each charging power class of electric vehicle is calculated, the actual charging time of AC 1 is

$$t_r = \frac{w}{P_1}. \quad (3)$$

The time of AC 2 is

$$t_r = \frac{w}{P_2}. \quad (4)$$

The time of DC is

$$t_r = \frac{w}{P_3}. \quad (5)$$

TABLE 1: Symbol descriptions.

Symbol	Significance
$\Delta A$	Power loss or power consumption in the circuit
$I$	Load current through the circuit
$R$	The resistance value of the wires that make up the circuit
$t$	The elapsed time for the circuit to be powered
$U$	Voltage

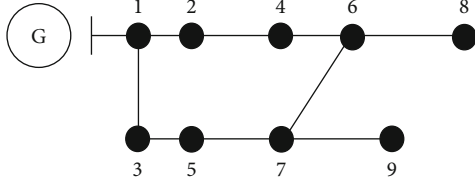


FIGURE 2: Distribution topology picture of a residential area.

TABLE 2: Minimum charging power ascending arrangement of 100 electric vehicles on Monday.

S/N	Charging time (h)	Monday		Classification
		Charging capacity (kWh)	Minimum power (kW)	
79	39.76	1.25	0.314	AC 1 10%
72	24.66	11.04	0.448	
...	...	...	...	
8	1.29	0.94	0.729	
39	10.09	7.64	0.758	
76	28.27	21.80	0.771	
42	0.83	0.68	0.819	
98	15.50	13.44	0.867	
44	9.17	8.07	0.880	
...	...	..	...	
36	1.47	2.45	1.667	
88	16.36	27.6	1.687	
23	10.18	17.31	1.700	
63	9.24	15.95	1.726	
66	13.62	24.36	1.789	
91	9.83	17.72	1.802	
12	6.88	12.5	1.817	
51	3.15	5.83	1.851	
37	7.9	14.83	1.877	DC 50%
...	...	..	...	
9	4.61	22.79	4.941	
16	3.08	15.95	5.179	
53	2.91	16.26	5.588	
11	1.61	9.96	6.186	
34	0.37	12.72	34.378	

Then, we can calculate that actual charging end time is

$$t_e = t_s + t_r. \quad (6)$$

Thus, the actual charging start-stop time of the electric vehicles is in  $[t_s, t_r]$ .

If we divide a 24-hour and 15-minute period into 96 time periods and use the cyclic algorithm to distribute the starting and stopping charging time data of 100 vehicles into 96 intervals, each interval is calculated by summing up the charging power of the corresponding AC-grade electric vehicles, and the charging power of 100 vehicles in 24 hours is obtained. By using MATLAB to plot the charging power of 100 vehicles in 24 hours, the charging load curve is plotted by function and the result is shown in Figure 3.

According to the difference between the highest power point and the lowest power point in the figure, the peak-valley difference of the charging load of 100 vehicles in 24 hours is 4200 kWh when the proportion of electric vehicles charged by AC 1 is 10%, charged by AC 2 is 40%, and charged by DC is 50%.

*3.3. Optimizing the Ratio of the Charging Power Level.* For 100 electric vehicles, the proportion of “AC 1,” “AC 2,” and “DC charging” electric vehicles after optimization is set to be  $x$ ,  $y$ , and  $z$ , respectively, to redistribute their charging power levels and adjust their actual charging time to achieve the goal of modernization. To reduce the equipment investment and reduce the peak-valley difference, we used the actual charging time and cycle algorithm to distribute time and accumulate power. The weight of the equipment investment and the peak-valley difference is 1 : 1, and the objective function is as follows:

$$A_{\min} = Y_{\min} + \Delta p. \quad (7)$$

In the process of optimizing the proportion of charging vehicles of different grades, we obtain the minimum objective function  $A_{\min}$  when the proportion of “AC 1,” “AC 2,” and “DC charging” electric vehicles is 46 : 53 : 1. Therefore, the minimum investment in equipment is

$$Y_{\min} = 140900000. \quad (8)$$

The minimum peak-valley difference is

$$\Delta p = W_{\max} - W_{\min} = 18810 \text{ kW}. \quad (9)$$

Using MATLAB, the charging load curve of 10,000 electric vehicles in 24 hours after optimization is drawn as shown in Figure 4. Compared with Figure 3, it is clear that the optimized peak value diminishes in magnitude and the peak-valley difference decreases significantly.

## 4. Optimizing Method

*4.1. Establishment of Model.* Based on the above analysis and calculation, we can obtain the total power consumption, which is shown as formulation (11).

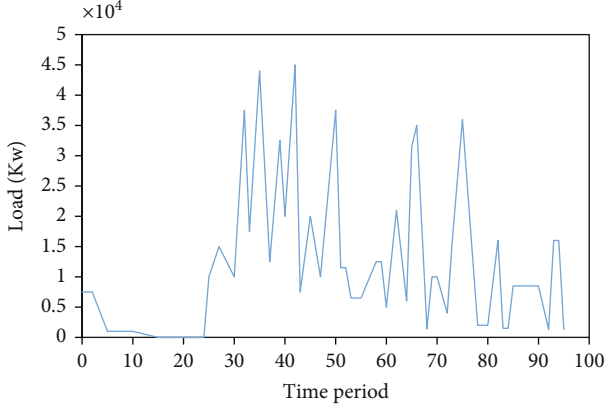


FIGURE 3: The charging load curve of 100 charged vehicles in 24 hours.

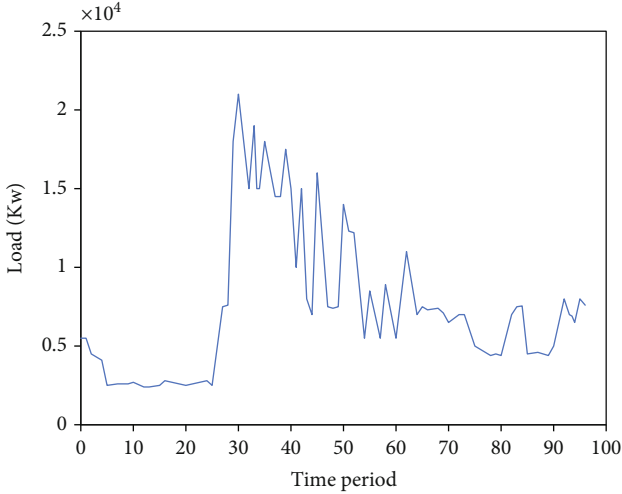


FIGURE 4: The charging load curve of 10,000 electric vehicles 24 hours after optimization.

The parallel parts of  $I$  and  $U$  produces Joule heat and active power loss, while the orthogonal part is the conversion of the electric field and magnetic field, which does not produce active power loss. When  $T = i$ , the voltage values of the nine nodes are  $U_{ji}$ ,  $i = 1, 2, 3, \dots, 9$ , respectively. When  $T = i$ , the number of two types of electric vehicles charged by charging piles is  $N_{1i}, N_{2i}$ , the number of two types of vehicles discharged by charging piles is  $N_{3i}, N_{4i}$ , the charge price is the CPI, and the selling price is the EPI.

The charging behavior optimization model of electric vehicles in a distribution network:

$$\begin{aligned}
 & \min W_{\text{extra}}(N) \\
 & \min S(N) \\
 & \text{s.t.} \begin{cases} U_i \geq 0.95 (i = 1, 2, 3, \dots, 9) \\ \frac{W_1}{\Delta t_3} < 5, \frac{W_1}{\Delta t_4} < 5 \\ \sum_{i=1}^{24} \sum_{j=1}^4 N_{ji} = 5000 \\ P_{c,j}(t) < 5, P_{c,j'}(t) < 5. \end{cases} \quad (10)
 \end{aligned}$$

Among them,  $i$  is the starting time of charging and discharging,  $j$  represents the AC 1 of electric vehicles, and  $j'$  represents the AC 2 of that.  $P_{c,j}(t)$  and  $P_{c,j'}(t)$ , respectively, indicate the charging power and discharging power of vehicle NO.J at time  $i$ .

$$\begin{aligned}
 \Delta P &= \frac{R * 10^{-3}}{U_b^2 \cos^2 \varphi} P^2, R = 0.16 * \frac{U_b^2}{P_B} = 25.644 \Omega, I = YV, Y \\
 &= G + B, V = ZI, Z = R + jX. \quad (11)
 \end{aligned}$$

4.2. *The Solution of Optimal Algorithms.* Taking  $N_{ji}$  as the optimization variable and making use of the genetic algorithm for multiobjective programming, we obtained the optimal peak and valley cost of the charging and discharging number about AC class 1 and AC class 2 in each period and the load curve of the power grid before and after optimization. The flow chart of the algorithm is shown in Figure 5.

The steps for solving the optimization algorithm are as follows:

- (Step 1) Initialize the load parameters and electric vehicle parameters of each node in the power grid, including the load value of the power grid, battery capacity, charging and discharging power, charging state, average energy consumption, and other parameters at each time of day
- (Step 2) Randomly generate peak and valley initial populations at each level and time, and then generate 200 initial populations within the range of devalues
- (Step 3) Compute the constraints of population, including equality constraints and inequality constraints. Populations that do not satisfy the constraints are discarded directly. The initial population that satisfies the constraints of the optimization model is obtained, and the discharge power and the number of EVs at all stages of charging and discharging are calculated
- (Step 4) Produce a progeny population by crossing and mutating in the survived individuals, under the circumstance where the crossover probability is  $P_c = 0.8$  and the mutation variable is  $P_m = 0.2$  [12]
- (Step 5) Optimize iteratively until the maximum genetic algebra equations have been reached, and calculate the discharge power of the electric vehicle and the number of electric vehicles at all stages of charging and discharging

4.3. *Example Analysis.* Taking Chengdu as an example, the parameters of electric vehicles are as follows in Table 3.

Under the guidance of changing electricity prices, some owners responded to the peak-valley price, charged and

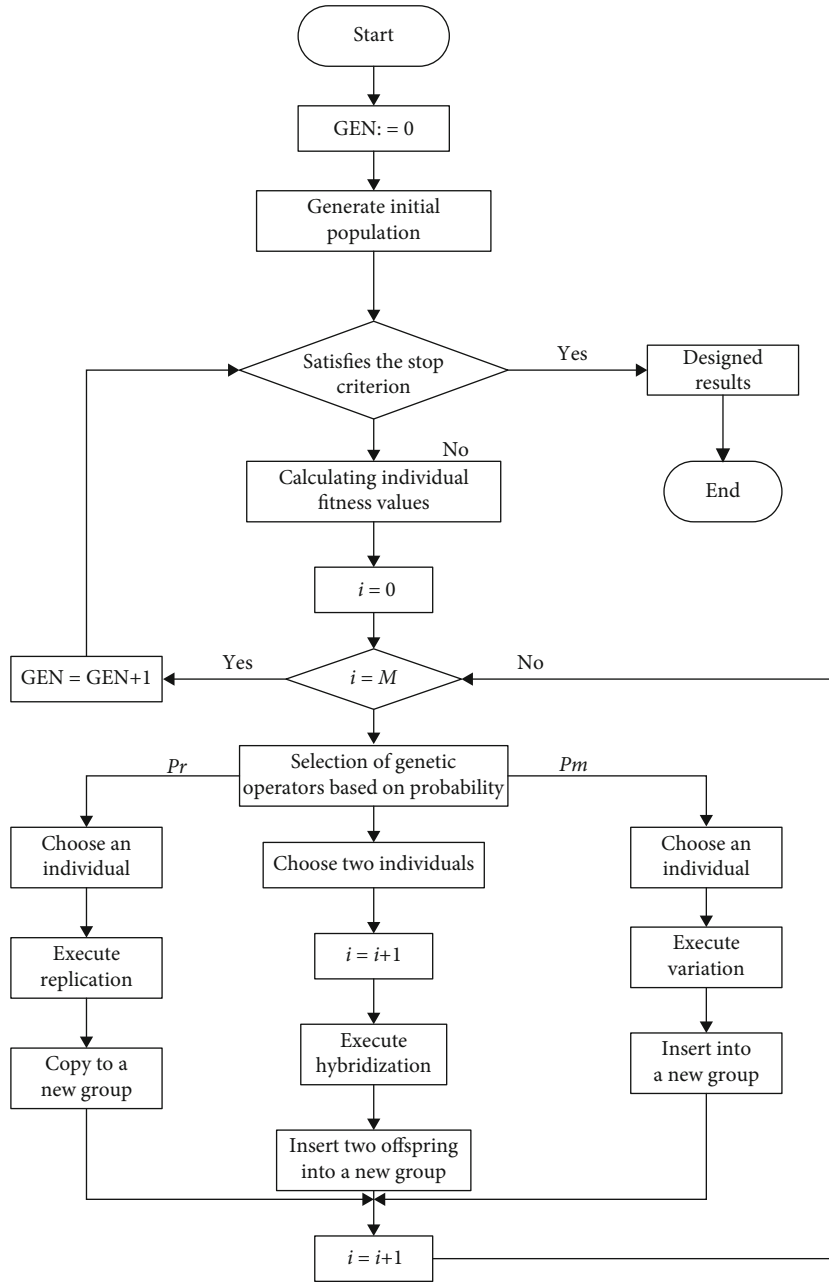


FIGURE 5: Optimal flow chart.

TABLE 3: Results.

Battery capacity	Energy consumption	Start date	Charged state
50 kWh	0.15 J	3.5	20%-90%

TABLE 4: Comparison of results before and after optimization.

Comparison	Peak load
Before optimization	2.7395 kWh
After optimization	2.587 kWh
Reducing peak load after optimization	0.0424 kWh

discharged according to the peak-valley price period, combined with the charging and discharging time. Finally, we found that orderly charging and discharging of electric vehicles indeed reduced the peak-valley difference and active power loss of the power grid. Based on the charging load curve of the Chengdu electric vehicles, the results before and after the peak-valley period optimization can be obtained. The results are shown in Table 4.

Thus, in the valley price stage, the charging load of EV had a valley filling effect, and in the peak price period, the discharge load of EV played a peak cutting role. Optimizing distribution networks can effectively reduce the peak-valley differences and achieve cost-saving goals.



## 5. Conclusion

New energy vehicles have far-reaching social significance regarding sustainable development, alleviating the energy crisis, improving the utilization rate for new energy, and helping the environment. The development path of the new energy automobile industry supported by the government policy is broad, and the future market application prospects are immeasurable. Realizing the scale and industrialization of new energy vehicles requires the improvement of core technology and product performance as well as regarding consumers' awareness of environmental protection [13]. However, there are also many problems in the interactions between electric vehicles and power grids.

The construction of new energy charging stations is expensive, and the disorderly charging of electric vehicles further increases the peak-valley gap of the power grid [14]. Therefore, it is necessary to establish a reasonable proportion model of electric vehicle charging power level, reduce the equipment investment for new energy charging stations, reduce the peak-valley difference of power grids, and reduce the power system consumption.

## Data Availability

To be frank, I derived the writing material from different journals as provided in the references. A MATLAB tool has been utilized to simulate our concept.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Authors' Contributions

Kaiyang Zhong is assigned to the conceptualization, methodology, software, validation, writing, editing, reviewing, and funding. Ping Wang is responsible formal analysis, investigation, and supervision. Jiaming Pei is responsible for the data curation, writing, editing, reviewing, and methodology. Jiyuan Xu did the conceptualization and investigation. Jiawen Xu participated in writing, editing, reviewing, and supervision.

## References

- [1] R. Wrobel and B. A. Mecrow, "Comprehensive review of additive manufacturing in construction of electrical machines," *IEEE Transactions on Energy*, vol. 35, no. 2, pp. 1054–1064, 2020.
- [2] G. A. Yue-fanga, L. I. Yong-shenga, L. U. Xinb, and R. E. Xian-yib, "Optimization control technique for wastewater treatment based on particle swarm algorithm," *Application Research of Computers*, vol. 28, no. 9, pp. 3318–3320, 2011.
- [3] Y. Xu, Y. Zheng, and Y. J. A. E. Yang, "On the movement simulations of electric vehicles: a behavioral model-based approach," *Applied Energy*, vol. 283, article 116356, 2021.
- [4] Y. Gao and S.-l. Xie, "Chaotic particle swarm optimization algorithm," *Computer Science*, vol. 31, no. 8, pp. 13–15, 2004.
- [5] Y. Miao, P. Hynan, A. von Jouanne, and A. J. E. Yokochi, "Current Li-ion battery technologies in electric vehicles and opportunities for advancements," *Energies*, vol. 12, no. 6, p. 1074, 2019.
- [6] Z. Li, A. Khajepour, and J. J. E. Song, "A comprehensive review of the key technologies for pure electric vehicles," *Energy*, vol. 182, pp. 824–839, 2019.
- [7] M. R. Khalid, M. S. Alam, A. Sarwar, and M. S. Jamil Asghar, "A Comprehensive review on electric vehicles charging infrastructures and their impacts on power-quality of the utility grid," *ETransportation*, vol. 1, article 100006, 2019.
- [8] T. Gnann, S. Funke, N. Jakobsson, P. Plötz, F. Sprei, and A. Bennehag, "Fast charging infrastructure for electric vehicles: today's situation and future needs," *Transportation Research Part D: Transport and Environment*, vol. 62, pp. 314–329, 2018.
- [9] J. Pei, "Solving the problem of Charging and Discharging of Electric Vehicles based on Particle Swarm Algorithm," *Book Solving the problem of Charging and Discharging of Electric Vehicles based on Particle Swarm Algorithm*, pp. 534–538, 2019.
- [10] A. Yan-ping, "Research on Intelligent Management System of New Energy Charging Makeup," *Book Research on Intelligent Management System of New Energy Charging Makeup*, 2016.
- [11] M. Shepero, J. Munkhammar, J. Widén, J. D. K. Bishop, and T. Boström, "Modeling of photovoltaic power generation and electric vehicles charging on city-scale: a review," *Renewable and Sustainable Energy Reviews*, vol. 89, pp. 61–71, 2018.
- [12] L. Kai-kai, "Overview of SOC research on pure electric vehicle power battery," *Internal combustion engine and accessories*, vol. 13, pp. 23–24, 2019.
- [13] T. Yue, L. Cong, H. Chang-Qiang, and Y. Wang, "Formation selection and optimization of multiple UAVs," *Computer simulation*, vol. 36, no. 5, pp. 98–104, 2019.
- [14] J. Perez-Vaquero, J. Landauer, H. Briesen, and P. Foerst, "A particle tracking velocimetry method to measure size and charge distributions in tribocharged powder particles," *Chemical Engineering Science*, vol. 229, article 116036, 2021.

## Research Article

# Recognition of Imbalanced Epileptic EEG Signals by a Graph-Based Extreme Learning Machine

Jie Zhou <sup>1</sup>, Xiongtao Zhang <sup>2</sup>, and Zhibin Jiang <sup>1</sup>

<sup>1</sup>Department of Computer Science and Engineering, Shaoxing University, Shaoxing, China

<sup>2</sup>School of Information Engineering, Huzhou University, Huzhou, China

Correspondence should be addressed to Jie Zhou; 799489588@qq.com

Received 8 April 2021; Revised 29 April 2021; Accepted 28 May 2021; Published 14 June 2021

Academic Editor: Shan Zhong

Copyright © 2021 Jie Zhou et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Epileptic EEG signal recognition is an important method for epilepsy detection. In essence, epileptic EEG signal recognition is a typical imbalanced classification task. However, traditional machine learning methods used for imbalanced epileptic EEG signal recognition face many challenges: (1) traditional machine learning methods often ignore the imbalance of epileptic EEG signals, which leads to misclassification of positive samples and may cause serious consequences and (2) the existing imbalanced classification methods ignore the interrelationship between samples, resulting in poor classification performance. To overcome these challenges, a graph-based extreme learning machine method (G-ELM) is proposed for imbalanced epileptic EEG signal recognition. The proposed method uses graph theory to construct a relationship graph of samples according to data distribution. Then, a model combining the relationship graph and ELM is constructed; it inherits the rapid learning and good generalization capabilities of ELM and improves the classification performance. Experiments on a real imbalanced epileptic EEG dataset demonstrated the effectiveness and applicability of the proposed method.

## 1. Introduction

Epilepsy is a common neurological disease that can cause recurrent seizures. During seizures, injury or life-threatening events may occur owing to the distraction or involuntary spasms of the patient [1, 2]. In the clinical diagnosis of various seizures, electroencephalogram (EEG) signal detection plays a crucial role [3]. This is because the epileptic brain releases characteristic waves during seizures. In recent years, an increasing number of machine learning-based methods have been applied for epileptic EEG signal recognition [4–8]. Figure 1 illustrates a machine learning method-based system for epileptic EEG signal recognition. The figure shows that an epileptic EEG signal recognition system involves the following three main steps: (1) a feature extraction method is used on original epileptic EEG signals for training and testing, (2) EEG signals after feature extraction for training are used to train the machine learning-based model to build an epileptic EEG signal recognition system, and (3) EEG signals after feature extraction for testing are then inputted into the epileptic EEG signal recognition system for detection.

Previously, many machine learning methods have been proposed for epileptic EEG signal recognition, such as the naive Bayes method (NB) [9], *K*-nearest neighbor (KNN) [10], support vector machine (SVM) [11], fuzzy system [12, 13], and extreme learning machine (ELM) [14, 15], and they have shown good effectiveness. In essence, epileptic EEG signal recognition is a typical imbalanced classification task [16, 17]. Compared with negative samples (people without epilepsy), positive samples (patients with epilepsy) have extremely low representation and cannot be well classified by traditional classifiers. Although the misclassification of positive samples has little effect on the model accuracy, it may cause serious medical malpractice. Therefore, traditional machine learning methods face several critical challenges for recognition of imbalanced epileptic EEG signals: (1) traditional machine learning methods often ignore the imbalance of epileptic EEG signals and misclassify positive samples, which may cause serious medical malpractice, and (2) existing imbalanced classification methods ignore the interrelationship between samples, resulting in poor classification performance. Therefore, building a classifier that considers

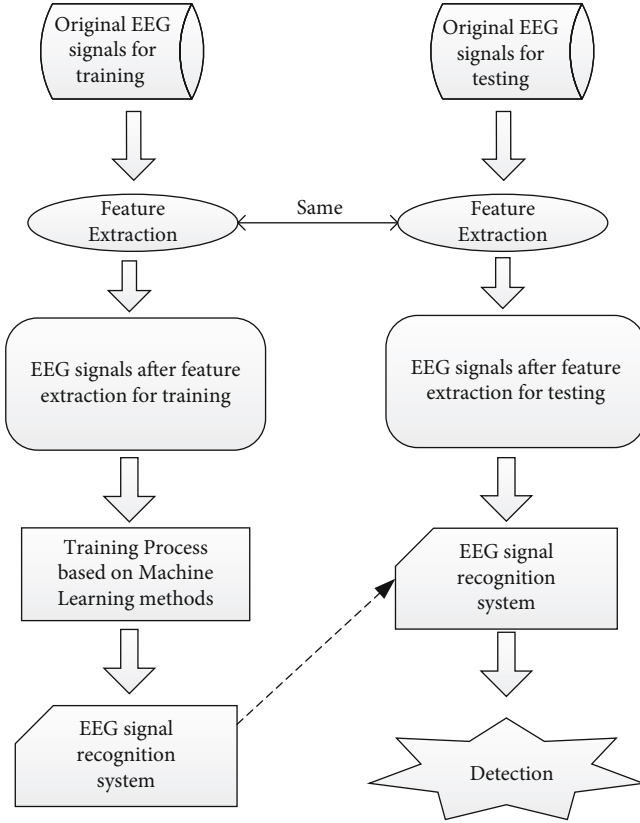


FIGURE 1: Illustration of the machine learning method-based system for epileptic EEG signal recognition.

the imbalance of the epileptic EEG signals and additional knowledge of samples becomes imperative for classification of imbalanced datasets with epileptic EEG signals.

To overcome these challenges, a novel imbalanced epileptic EEG signal recognition method based on a graph and ELM is proposed in this study. ELM has become a classical machine learning method with its solid theoretical foundations, fast training speed, and good predictive performance [18, 19]. Although ELM can universally approximate to any continuous functions, it is not effective for classifying imbalanced datasets. Therefore, it is necessary to adopt strategies to make ELM correctly classify positive samples to obtain a reasonable classification result of an imbalanced dataset. Previously, numerous imbalanced ELM-based methods have been proposed. For example, Zong et al. [20] proposed the weighted extreme learning machine (WELM), which pioneered the application of ELM in imbalanced classification. Similarly, Zhang and Ji [21] proposed a fuzzy ELM (FELM), which regulated the distributions of penalty factors by inserting a fuzzy matrix. Yu et al. [22] proposed a special cost-sensitive ELM (ODOC-ELM) for imbalanced classification problems. Li et al. [23] proposed an ensemble WELM algorithm based on the AdaBoost framework to learn the weights of different samples adaptively. Yang et al. [24] proposed a novel ELM-based imbalanced classification method by estimating the probability density distributions for two imbalanced classes. Shukla and Yadav [25] combined CC-ELM with WELM to propose a regularized weighted CC-ELM.

Xiao et al. [26] proposed an imbalanced ELM-based algorithm for two classes of classification tasks by solving each class classification error. Du et al. [27] proposed an online sequential extreme learning machine with under- and over-sampling (OSELM-UO) for online imbalanced big data classification. In addition, some ELM-based imbalanced methods, such as ensemble weighted ELM [28], class-specific cost regulation ELM [29], label-weighted extreme learning machine [30], and class-specific ELM [31], have also been proposed. However, to the best of our knowledge, there is no study that uses imbalanced ELM methods for epileptic EEG signal recognition; therefore, it is necessary to propose such a method for epileptic EEG signal recognition.

In this study, inspired by WELM, we propose a novel graph-based ELM (G-ELM) for imbalanced epileptic EEG signal recognition. First, we use the graph theory to construct a relationship graph of samples according to their data distribution. Then, we combine the relationship graph with ELM to propose G-ELM. The experimental results on a real imbalanced epileptic EEG dataset show that the proposed method can address imbalanced classification of epileptic EEG signals effectively. The main contributions of this study are as follows.

- (1) The proposed G-ELM sets the compensation for loss of positive samples to be greater than that of negative samples based on graph theory and then combines with the ELM to classify imbalanced data effectively. It is a novel imbalanced ELM-based method, which attains a good classification performance and inherits the rapid learning and good generalization capabilities of ELM
- (2) The proposed imbalanced classification method attempts to consider both the imbalance and interrelationship of epileptic EEG samples to obtain better performance for imbalanced epileptic EEG signal recognition. It can be utilized for imbalanced epileptic EEG signal recognition. It not only realizes effective classification of imbalanced epileptic EEG signals from a new perspective but also expands application of ELM-based algorithms
- (3) We use six imbalanced classification evaluation indices, i.e., accuracy, precision, recall,  $F$ -measure,  $G$ -means, and AUC, to compare the performance of the proposed G-ELM and the existing imbalanced ELM-based methods. Extensive experiments on a real imbalanced epileptic EEG dataset indicate that the proposed method can address imbalanced epileptic EEG signal recognition effectively and outperform the existing imbalanced ELM-based methods

The rest of this paper is organized as follows. Section 2 introduces the background underlying the proposed epileptic EEG recognition method. In Section 3, the details of the proposed G-ELM are presented. The performance of the proposed method is evaluated with several comparative methods in Section 4. The conclusions of this paper are provided in Section 5.

TABLE 1: Descriptions of the five groups in *Bonn*.

	Groups	Sizes of groups	Descriptions of EEG signals
<i>Healthy people</i>	A	100	EEG signals obtained with eyes open
	B	100	EEG signals obtained with eyes closed
	C	100	EEG signals measured in the hippocampus of the brain during seizure-free intervals
<i>Epileptic people</i>	D	100	EEG signals measured in the epileptogenic zone during seizure-free intervals
	E	100	EEG signals measured during seizure activity

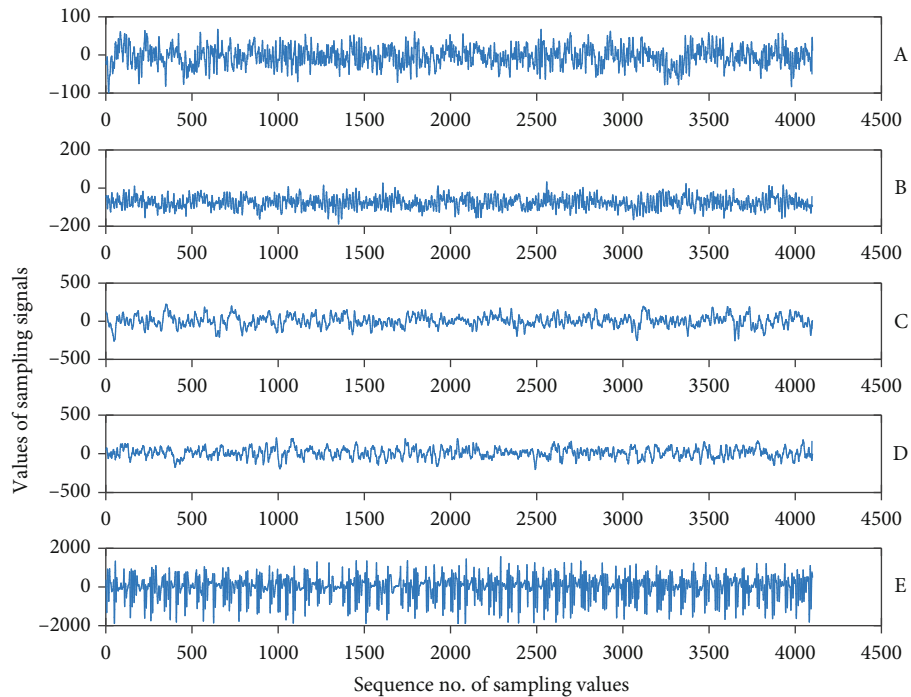


FIGURE 2: Representative original epileptic EEG signals of five different groups.

## 2. Background

In this section, we briefly describe the background related to the proposed epileptic EEG signal recognition method. It includes the epileptic EEG dataset, the feature extraction methods, and the classical ELM, which are used for epileptic EEG signal detection.

**2.1. Epileptic EEG Dataset.** The real epileptic EEG dataset used in this paper is *Bonn* [32], which is from the University of Bonn, Germany. It can be publicly downloaded from the following website (<http://www.epileptologie-bonn.de/cms/upload/workgroup/lehnertz/eegdata.html>). There are five groups (denoted by A–E, respectively) in *Bonn*. In each group, there are 100 samples of 23.6s segments. Detailed descriptions of the five groups are given in Table 1. Groups A and B are segments acquired from five healthy volunteers with eyes open (Group A) and eyes closed (Group B). Groups C–E are segments acquired from volunteers with epilepsy. In Group C, EEG signals are measured in the hippocampus of the brain during seizure-free intervals and those in Group D are measured in the epileptogenic zone during seizure-free intervals. In Group E, EEG signals are measured during

seizure activity. Five representative original epileptic EEG signals of five different groups are shown in Figure 2.

**2.2. Feature Extraction.** Many studies [33–35] have shown that the original EEG signals cannot be directly used for training machine learning-based models and that feature extraction is a necessary step. This is because the original EEG signals are usually high dimensional, stochastic, non-stationary, and nonlinear and the background noise in the original signals is very complex. The commonly used feature extraction methods can be divided into three main categories: time domain analysis, frequency domain analysis, and time-frequency analysis. Time domain analysis-based methods extract the features by analyzing the characteristics of original EEG signals, such as mean, variance, amplitude, and kurtosis [36]. Frequency domain analysis-based methods usually analyze the EEG signals in the frequency domain to extract the features, such as fast Fourier transforms [37] and short-time Fourier transforms [38]. As for time-frequency analysis methods, the information of time and frequency domain is considered simultaneously to extract the features from original epileptic EEG signals. Typical time-frequency analysis-based methods are wavelet transform

methods [39, 40]. In this paper, we use the wavelet packet decomposition [40] for feature extraction from original epileptic EEG signals to simultaneously utilize the information of time and frequency domain.

**2.3. ELM.** ELM [19], which was first proposed by Huang et al., is a single-hidden-layer feedforward neural network [41]. It can directly optimize the output weight of the hidden layer by setting the number of hidden nodes, without paying attention to the weight and offset of the input layer, which can be generated randomly. Compared with other traditional supervised learning methods, it has good generalization ability and high learning speed. Figure 3 shows the network structure of an ELM.

ELM considers both empirical and structural risks, and its objective function is as follows:

$$\begin{aligned} \min \quad & \frac{1}{2} \|\boldsymbol{\beta}\|^2 + \frac{1}{2} C \sum_{i=1}^n \|\varepsilon_i\|^2, \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{x}_i)\boldsymbol{\beta} = y_i + \varepsilon_i, \quad i = 1, 2, \dots, n, \end{aligned} \quad (1)$$

where

$$\mathbf{H}_{n \times m} = (\mathbf{h}(\mathbf{x}_1), \mathbf{h}(\mathbf{x}_2), \dots, \mathbf{h}(\mathbf{x}_n))^T = \begin{pmatrix} h_1(\mathbf{x}_1) & \dots & h_m(\mathbf{x}_1) \\ \vdots & \ddots & \vdots \\ h_1(\mathbf{x}_n) & \dots & h_m(\mathbf{x}_n) \end{pmatrix} \quad (2)$$

represents the hidden layer feature matrix, where  $h_i(\mathbf{x}_j) = g(\mathbf{A}^{(i)}\mathbf{x}_j + b_i)$ ,  $\mathbf{A}^{(i)}$  represents the  $i$ th row of the weight matrix  $\mathbf{A}_{m \times d}$ ,  $\mathbf{b} = (b_1, b_2, \dots, b_m)^T$  represents the offset,  $\mathbf{x}_i \in \mathbb{R}^d$ ,  $i = 1, 2, \dots, n$  denotes the training samples,  $n$  is the number of training samples,  $d$  is dimension, and  $m$  is the number of hidden nodes;  $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^T$  is the error matrix between the network outputs and the target outputs.  $C$  is a penalty parameter, which can adjust the accuracy and generalization ability of the ELM.

The optimization problem in (1) can be solved based on the Karush–Kuhn–Tucker theory. The output weight of ELM can be calculated by

$$\boldsymbol{\beta} = \begin{cases} \mathbf{H}^T \left( \frac{\mathbf{I}}{C} + \mathbf{H}\mathbf{H}^T \right)^{-1} \mathbf{Y} & n < m, \\ \left( \frac{\mathbf{I}}{C} + \mathbf{H}\mathbf{H}^T \right)^{-1} \mathbf{H}^T \mathbf{Y} & n \geq m. \end{cases} \quad (3)$$

### 3. Graph-Based Extreme Learning Machine

In this section, a graph-based ELM (G-ELM) is proposed. We first introduce the relationship graph of an imbalanced dataset and then develop the proposed imbalanced classification method G-ELM by combining the relationship graph with an ELM.

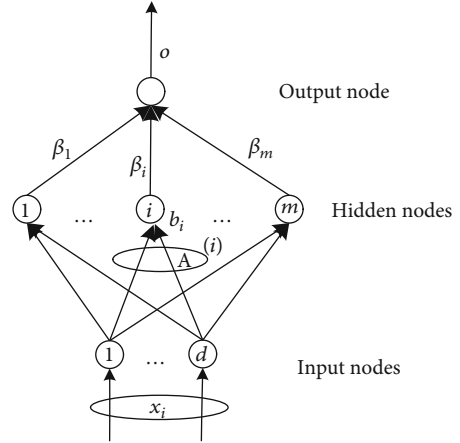


FIGURE 3: ELM network structure.

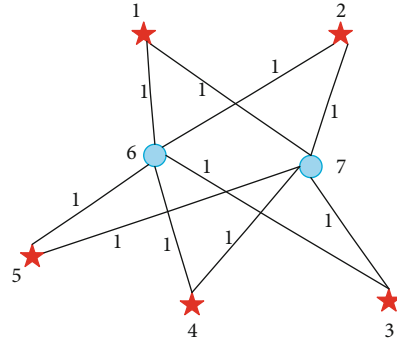


FIGURE 4: Graph structure of an imbalanced synthetic dataset.

**3.1. Relationship Graph of an Imbalanced Dataset.** In the context of imbalanced classification problem, the relationships between the training samples can be regarded as an undirected graph.

Undirected graph can be expressed as  $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ , where  $\mathbf{V}$  is the vertex set of graph  $\mathbf{G}$  and  $\mathbf{E}$  is the edge set of graph  $\mathbf{G}$ . Figure 4 shows an undirected graph of an imbalanced synthetic dataset with 7 samples, where 2 positive samples are represented by a *blue circle* and 5 negative samples are represented by a *red star*. All samples are numbered for subsequent display. Note that there are connections between samples in different classes and the weight is 1. Samples in the same class are not connected.

The elements of an adjacency matrix  $\mathbf{W}$  can be defined as follows:

$$w_{ij} = \begin{cases} 1, & y_i \neq y_j, \\ 0, & y_i = y_j. \end{cases} \quad (4)$$

Here,  $y_i \in \mathbf{Y}$  is the label of  $\mathbf{x}_i$ .

According to the above definition of the adjacency matrix  $\mathbf{W}$ , we can see that the distance of the samples in the same class can be considered 0. For samples in different classes, the distance between them can be considered 1.

Then, the relationship graph matrix can be expressed as

$$\mathbf{L} = \mathbf{D} - \mathbf{W}, \quad (5)$$

where  $\mathbf{D} = \text{diag}(\mathbf{W} \cdot \mathbf{1}_{n \times 1})$  is the degree matrix;  $\mathbf{1}_{n \times 1}$  stands for a vector with  $n \times 1$ , whose elements are exactly 1;  $n$  is the number of training samples.

As for the imbalanced dataset  $\mathbf{X}$ , we need to increase the loss of misclassification of positive samples because the misclassification of positive samples (patients with epilepsy) could cause serious consequences. This can be realized by regulating the degree matrix  $\mathbf{D}$ . The shortcomings of the cost learning algorithm can be compensated by increasing the relationship between samples. Therefore, the relationship graph not only ensures the accuracy of positive sample classification but also makes up for the lack of the mutual relationships and prior knowledge between samples.

According to the above description, the relationship graph matrix  $\mathbf{L}$  of the synthetic dataset in Figure 4 can be expressed as

$$\mathbf{L} = 4 \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{matrix} & \begin{bmatrix} 2 & 0 & 0 & 0 & 0 & -1 & -1 \\ 0 & 2 & 0 & 0 & 0 & -1 & -1 \\ 0 & 0 & 2 & 0 & 0 & -1 & -1 \\ 0 & 0 & 0 & 2 & 0 & -1 & -1 \\ 0 & 0 & 0 & 0 & 2 & -1 & -1 \\ -1 & -1 & -1 & -1 & -1 & 5 & 0 \\ -1 & -1 & -1 & -1 & -1 & 0 & 5 \end{bmatrix} \end{matrix} \quad (6)$$

**3.2. Objective Function of G-ELM.** According to the above relationship graph and ELM, the objective function of the G-ELM can be expressed as follows:

$$\begin{aligned} \min \quad & \frac{1}{2} \|\boldsymbol{\beta}\|^2 + \frac{1}{2} C \sum_{i=1}^n \sum_{j=1}^n \varepsilon_i L_{ij} \varepsilon_j, \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{x}_i) \boldsymbol{\beta} = y_i - \varepsilon_i, \quad i = 1, 2, \dots, n. \end{aligned} \quad (7)$$

Here,  $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathbb{R}^{d \times n}$ ,  $n$  is the number of samples in  $\mathbf{X}$ ,  $d$  is the sample dimension, and  $\mathbf{Y} = [y_1, y_2, \dots, y_n]^T$  represents the true class label of the samples.  $\mathbf{H}$  and  $\mathbf{h}(\mathbf{x}_i)$  are the same as defined in ELM.  $\boldsymbol{\beta} = [\beta_1, \beta_2, \dots, \beta_m]^T$  represents the output weight vector.  $\boldsymbol{\varepsilon} = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n]^T$  represents the loss between the network outputs and the target outputs.

$$\mathbf{L} = \begin{bmatrix} L_{11} & \cdots & L_{1n} \\ \vdots & \ddots & \vdots \\ L_{n1} & \cdots & L_{nn} \end{bmatrix}_{n \times n}. \quad (8)$$

Equation (8) is the relationship graph matrix of the samples.

By comparing (7) with (1), we can see that G-ELM is an improved version of ELM and still has the characteris-

tics of high learning speed and strong generalization ability from ELM.

**3.3. Solution of G-ELM.** In this subsection, we attempt to optimize the objective function of G-ELM. According to [20], the objective function of G-ELM is a convex optimization problem. The specific optimization solution process is as follows:

The Lagrangian function corresponding to (7) is

$$J = \frac{1}{2} \|\boldsymbol{\beta}\|^2 + \frac{1}{2} C \sum_{i=1}^n \sum_{j=1}^n \varepsilon_i L_{ij} \varepsilon_j - \sum_{i=1}^n \alpha_i (\mathbf{h}(\mathbf{x}_i) \boldsymbol{\beta} - y_i + \varepsilon_i). \quad (9)$$

Let the derivation of  $J$  with respect to  $\boldsymbol{\beta}$ ,  $\varepsilon_i$ ,  $\alpha_i$  equal to zero:

$$\frac{\partial J}{\partial \boldsymbol{\beta}} = \boldsymbol{\beta} - \sum_{i=1}^n \alpha_i \mathbf{h}(\mathbf{x}_i) = 0 \Rightarrow \boldsymbol{\beta} = \mathbf{H}^T \boldsymbol{\alpha}, \quad (10a)$$

$$\frac{\partial J}{\partial \varepsilon_i} = -\alpha_i + C \sum_{j=1}^n L_{ij} \varepsilon_j = 0 \Rightarrow \boldsymbol{\alpha} = \mathbf{C} \mathbf{L} \boldsymbol{\varepsilon}, \quad (10b)$$

$$\frac{\partial J}{\partial \alpha_i} = \mathbf{h}(\mathbf{x}_i) \boldsymbol{\beta} - y_i + \varepsilon_i = 0 \Rightarrow \mathbf{H} \boldsymbol{\beta} - \mathbf{Y} + \boldsymbol{\varepsilon} = 0. \quad (10c)$$

Substituting (10a) and (10b) into (10c), we obtain

$$\begin{aligned} \mathbf{H} \mathbf{H}^T \boldsymbol{\alpha} - \mathbf{Y} + \frac{1}{C} \mathbf{L}^{-1} \boldsymbol{\alpha} &= 0 \Rightarrow \left( \mathbf{H} \mathbf{H}^T + \frac{1}{C} \mathbf{L}^{-1} \right) \boldsymbol{\alpha} = \mathbf{Y} \\ &\Rightarrow \left( \mathbf{L} \mathbf{H} \mathbf{H}^T + \frac{\mathbf{I}}{C} \right) \boldsymbol{\alpha} = \mathbf{L} \mathbf{Y}. \end{aligned} \quad (11)$$

Combining (10a) and (11), we obtain

$$\boldsymbol{\beta} = \begin{cases} \mathbf{H}^T \left( \mathbf{L} \mathbf{H} \mathbf{H}^T + \frac{\mathbf{I}}{C} \right)^{-1} \mathbf{L} \mathbf{Y} & n < m, \\ \left( \mathbf{H}^T \mathbf{L} \mathbf{H} + \frac{\mathbf{I}}{C} \right)^{-1} \mathbf{H}^T \mathbf{L} \mathbf{Y} & n \geq m. \end{cases} \quad (12)$$

With the obtained solution, i.e.,  $\boldsymbol{\beta}^*$ , the predicted class label of the testing sample can be obtained as follows:

$$\mathbf{y}_{\text{test}} = \text{sign}(\mathbf{x}_{\text{test}} \boldsymbol{\beta}^*), \quad (13)$$

where  $\mathbf{x}_{\text{test}}$  is a testing sample.

**3.4. Learning Algorithm of G-ELM.** According to the above derivation, the implementation of G-ELM is summarized in Algorithm 1.

## 4. Experiments

To demonstrate the effectiveness of the proposed G-ELM, we conducted extensive experiments on a real epileptic EEG dataset. The proposed G-ELM was verified by comparing it with five ELM-based methods, i.e., ELM [19], W1-ELM [20], W2-ELM [20], R1-ELM [25], and R2-ELM [25], using

**Input:** The training samples  $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathbb{R}^{d \times n}$  and their corresponding labels  $\mathbf{Y} = [y_1, y_2, \dots, y_n]^T$ , where  $\mathbf{x}_i \in \mathbb{R}^d (i = 1, 2, \dots, n)$ . The number of the hidden nodes  $m$ ; the input weights  $\mathbf{A} \in \mathbb{R}^{m \times d}$  and input biases  $\mathbf{b} \in \mathbb{R}^m$ ; the penalty parameter  $C$ .

**Output:** The predicted class label of the testing sample  $\mathbf{x}_{test}$ .

**Step 1:** Construct the mapping matrix of hidden layer  $\mathbf{H}$  according to Eq. (2).

**Step 2:** Compute the relationship graph matrix corresponding to the training samples  $\mathbf{X}$  according to Eq. (4) and Eq. (5).

**Step 3:**  
**If**  $n < m$   
**Then** compute the output weight vector  $\boldsymbol{\beta}^*$  using the first formula in Eq. (12).  
**Else** compute  $\boldsymbol{\beta}^*$  using the second formula in Eq. (12).

**Step 4:** Return the predicted class label of the testing sample  $\mathbf{y}_{test} = \text{sign}(\mathbf{x}_{test} \boldsymbol{\beta}^*)$ .

ALGORITHM 1: G-ELM.

TABLE 2: Nine datasets constructed from *Bonn* for the experiments.

Datasets	Negative classes	Positive class	Sizes	IR
D1	A, C	E	300	2
D2	B, C	E	300	2
D3	A, D	E	300	2
D4	B, D	E	300	2
D5	A, B, C	E	400	3
D6	A, B, D	E	400	3
D7	A, C, D	E	400	3
D8	B, C, D	E	400	3
D9	A, B, C, D	E	500	4

six imbalanced classification evaluation indices and average standard deviation on the real *Bonn* dataset. Except for ELM, the other comparison methods are imbalanced classification methods. All the experiments were conducted on a computer with Intel Core i5-3317U 1.70 GHz CPU and 16 GB RAM by using MATLAB 2016a. The details of the experimental settings and results are presented in the following sections.

**4.1. Data Preparation.** Although the real *Bonn* dataset has been used in many studies, the way of using it in this study differs from those in previous works. To evaluate the performance of the proposed G-ELM, nine imbalanced datasets were generated from the original five groups of EEG signals to simulate the imbalanced classification scenario. The details of the nine datasets are summarized in Table 2. In each dataset, the EEG signals of patients with epilepsy (E) were regarded as a positive class, while the other groups were regarded as a negative class, to identify whether the patients with epilepsy are experiencing seizure activity. A brief description of the five groups (A, B, C, D, and E) can be found in Table 1. The last column of Table 2 is IR, which is used to show the degree of imbalance of the dataset. IR can be defined as follows:

$$\text{IR} = \frac{n^-}{n^+}, \quad (14)$$

where  $n_+$  and  $n_-$  represent the number of samples of the positive class and the negative class, respectively.

In our experiment, we randomly partitioned each dataset. In each dataset, 80% of the dataset were used for training and the remaining 20% were used for testing.

**4.2. Evaluation Indices.** In our experiments, we used six imbalanced classification evaluation indices to evaluate all the adopted methods. The six imbalanced classification evaluation indices were accuracy, precision, recall, *F*-measure, *G*\_means, and AUC, which can be, respectively, defined as

$$\text{Accuracy} = \frac{(\text{TN} + \text{TP})}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})},$$

$$\text{Precision} = \frac{\text{TP}}{(\text{TP} + \text{FP})},$$

$$\text{Recall} = \frac{\text{TP}}{(\text{TP} + \text{FN})},$$

$$F\text{-measure} = \frac{(2 \cdot \text{recall} \cdot \text{precision})}{(\text{recall} + \text{precision})},$$

$$G\text{-means} = \sqrt{\frac{\text{TP}}{(\text{TP} + \text{FN})} \cdot \frac{\text{TN}}{(\text{TN} + \text{FP})}}. \quad (15)$$

Here, TP is the number of true positive samples, FN is the number of false negative samples, TN is the number of true negative samples, and FP is the number of false positive samples, respectively.

$$\text{AUC} = \sum_{i \in \mathcal{N}_+} \sum_{j \in \mathcal{N}_-} \frac{I(P(\mathbf{x}_i), P(\mathbf{x}_j))}{n_+ n_-}, \quad (16)$$

where  $\mathcal{N}_+$  is the set of all the indexes of the positive samples and  $\mathcal{N}_-$  is the set of those of the negative samples;  $n_+ = |\mathcal{N}_+|$  and  $n_- = |\mathcal{N}_-|$ .  $P(\mathbf{x})$  is the prediction value of  $\mathbf{x}$ .  $I(\cdot)$  is the indicator function

$$I(P(\mathbf{x}_i), P(\mathbf{x}_j)) = \begin{cases} 1, & P(\mathbf{x}_i) > P(\mathbf{x}_j), \\ 0.5, & P(\mathbf{x}_i) = P(\mathbf{x}_j), \\ 0, & P(\mathbf{x}_i) < P(\mathbf{x}_j). \end{cases} \quad (17)$$

TABLE 3: Results of all methods on the accuracy index.

Datasets	ELM	W1-ELM	W2-ELM	R1-ELM	R2-ELM	G-ELM
D1	0.8903(0.0027)	0.8593(0.0149)	0.8815(0.0176)	0.9389(0.0012)	0.9178(0.0041)	0.9489(0.0011)
D2	0.8744(0.0033)	0.8513(0.0150)	0.8574(0.0139)	0.9333(0.0021)	0.9207(0.0025)	0.9533(0.0023)
D3	0.8867(0.0056)	0.8472(0.0108)	0.8513(0.0125)	0.9067(0.0028)	0.8989(0.0022)	0.9511(0.0011)
D4	0.8633(0.0052)	0.8593(0.0153)	0.8452(0.0179)	0.9167(0.0019)	0.8989(0.0024)	0.9433(0.0024)
D5	0.7475(0.0081)	0.8594(0.0144)	0.8674(0.0229)	0.8917(0.0034)	0.7625(0.0058)	0.9383(0.0035)
D6	0.8325(0.0097)	0.8712(0.0167)	0.8657(0.0201)	0.9217(0.0031)	0.8258(0.0109)	0.9425(0.0028)
D7	0.8783(0.0035)	0.8774(0.0192)	0.8681(0.0088)	0.9283(0.0020)	0.9158(0.0023)	0.9550(0.0018)
D8	0.9042(0.0045)	0.8519(0.0166)	0.8625(0.0128)	0.9225(0.0036)	0.9125(0.0036)	0.9533(0.0018)
D9	0.9102(0.0041)	0.8675(0.0151)	0.8658(0.0127)	0.9180(0.0039)	0.8880(0.0029)	0.9480(0.0033)
Ave	0.8653	0.8605	0.8628	0.9198	0.8823	0.9482
Ave. std	0.0052	0.0153	0.0155	0.0027	0.0041	0.0022

TABLE 4: Results of all methods on the recall index.

Datasets	ELM	W1-ELM	W2-ELM	R1-ELM	R2-ELM	G-ELM
D1	0.7467(0.0072)	0.8231(0.0194)	0.8048(0.0219)	0.8533(0.0056)	0.8967(0.0038)	0.9367(0.0021)
D2	0.7233(0.0083)	0.8051(0.0351)	0.7881(0.0194)	0.8467(0.0034)	0.8667(0.0031)	0.9033(0.0038)
D3	0.8403(0.0088)	0.7933(0.0172)	0.7947(0.0184)	0.9233(0.0045)	0.9233(0.0055)	0.9533(0.0034)
D4	0.7867(0.0064)	0.7625(0.0199)	0.8156(0.0354)	0.8767(0.0045)	0.8833(0.0047)	0.9196(0.0034)
D5	0.7167(0.0110)	0.8006(0.0201)	0.8509(0.0218)	0.7967(0.0080)	0.8567(0.0045)	0.8489(0.0035)
D6	0.6867(0.0134)	0.8037(0.0270)	0.8374(0.0166)	0.8333(0.0054)	0.8498(0.0065)	0.8805(0.0056)
D7	0.7405(0.0093)	0.7825(0.0273)	0.7751(0.0162)	0.8597(0.0042)	0.8906(0.0045)	0.8933(0.0028)
D8	0.7667(0.0054)	0.7619(0.0254)	0.8125(0.0272)	0.8933(0.0053)	0.9033(0.0058)	0.9267(0.0028)
D9	0.6902(0.0114)	0.7972(0.0150)	0.8143(0.0172)	0.8533(0.0034)	0.8667(0.0054)	0.8896(0.0045)
Ave	0.7442	0.7922	0.8104	0.8596	0.8819	0.9058
Ave. std	0.0090	0.0229	0.0216	0.0049	0.0049	0.0035

TABLE 5: Results of all methods on the precision index.

Datasets	ELM	W1-ELM	W2-ELM	R1-ELM	R2-ELM	G-ELM
D1	0.9078(0.0058)	0.8247(0.0304)	0.8342(0.0301)	0.9619(0.0087)	0.8648(0.0097)	0.9126(0.0029)
D2	0.8801(0.0077)	0.7878(0.0313)	0.7901(0.0298)	0.9497(0.0077)	0.8914(0.0063)	0.9569(0.0091)
D3	0.8279(0.0125)	0.8009(0.0261)	0.7911(0.0212)	0.8208(0.0062)	0.8040(0.0053)	0.9053(0.0026)
D4	0.8054(0.0124)	0.8725(0.0375)	0.7729(0.0290)	0.8761(0.0079)	0.8267(0.0060)	0.9131(0.0085)
D5	0.5031(0.0107)	0.8438(0.0360)	0.8609(0.0434)	0.7845(0.0147)	0.5178(0.0078)	0.9007(0.0102)
D6	0.6825(0.0243)	0.8833(0.0388)	0.8502(0.0399)	0.8541(0.0112)	0.6244(0.0216)	0.8952(0.0125)
D7	0.7714(0.0131)	0.7905(0.0400)	0.8817(0.0349)	0.8568(0.0088)	0.7985(0.0067)	0.9266(0.0092)
D8	0.8431(0.0149)	0.8466(0.0359)	0.7732(0.0233)	0.8194(0.0120)	0.7860(0.0119)	0.8932(0.0083)
D9	0.8374(0.0176)	0.8414(0.0204)	0.8229(0.0367)	0.7701(0.0139)	0.6733(0.0087)	0.8650(0.0174)
Ave	0.7843	0.8324	0.8197	0.8548	0.7541	0.9076
Ave. std	0.0132	0.0329	0.0320	0.0101	0.0093	0.0090

$F$ -measure,  $G$ \_means, and AUC are commonly used imbalanced classification evaluation indices [42–45].

**4.3. Adopted Methods and Parameter Settings.** In the experiments, five ELM-based methods, i.e., ELM [19], W1-ELM [20], W2-ELM [20], R1-ELM [25], and R2-ELM [25], were adopted for comparisons with G-ELM. Referring to the guidelines in [2, 20, 46], a grid search strategy based

on  $G$ \_means was used to determine appropriate parameters of all the methods. We set parameter  $C$  in the range of  $2^{[-28:28]}$  and parameter  $m$  in the range of  $\{50, 100, 300, 500, 1000\}$  for all the adopted methods. All the adopted methods were run ten times on each generated imbalanced dataset. The average experimental results corresponding to the six imbalanced classification evaluation indices are reported.



TABLE 6: Results of all methods on the  $F$ -measure index.

Datasets	ELM	W1-ELM	W2-ELM	R1-ELM	R2-ELM	G-ELM
D1	0.8187(0.0047)	0.7716(0.0111)	0.8103(0.0218)	0.9031(0.0013)	0.8797(0.0053)	0.9244(0.0016)
D2	0.7931(0.0057)	0.7561(0.0225)	0.7808(0.0151)	0.8946(0.0028)	0.8786(0.0035)	0.9285(0.0031)
D3	0.8323(0.0075)	0.7620(0.0130)	0.7680(0.0151)	0.8686(0.0036)	0.8590(0.0028)	0.9285(0.0017)
D4	0.7942(0.0062)	0.7628(0.0152)	0.7539(0.0220)	0.8755(0.0021)	0.8536(0.0033)	0.9158(0.0032)
D5	0.5883(0.0073)	0.7463(0.0123)	0.7636(0.0188)	0.7872(0.0044)	0.6444(0.0054)	0.8739(0.0053)
D6	0.6762(0.0121)	0.8032(0.0250)	0.8285(0.0231)	0.8423(0.0052)	0.7148(0.0134)	0.8854(0.0040)
D7	0.7528(0.0061)	0.7924(0.0261)	0.7507(0.0116)	0.8575(0.0033)	0.8412(0.0037)	0.9089(0.0032)
D8	0.8012(0.0074)	0.7708(0.0233)	0.7797(0.0176)	0.8531(0.0056)	0.8387(0.0053)	0.9089(0.0030)
D9	0.7545(0.0108)	0.7130(0.0178)	0.7069(0.0149)	0.8079(0.0074)	0.7566(0.0044)	0.8744(0.0068)
Ave	0.7568	0.7642	0.7714	0.8544	0.8074	0.9054
Ave. std	0.0075	0.0185	0.0178	0.0040	0.0052	0.0036

TABLE 7: Results of all methods on the  $G$ \_means index.

Datasets	ELM	W1-ELM	W2-ELM	R1-ELM	R2-ELM	G-ELM
D1	0.8471(0.0040)	0.8242(0.0102)	0.8516(0.0174)	0.9150(0.0012)	0.9122(0.0033)	0.9458(0.0012)
D2	0.8286(0.0047)	0.8318(0.0197)	0.8402(0.0125)	0.9092(0.0016)	0.9057(0.0023)	0.9399(0.0014)
D3	0.8738(0.0055)	0.8213(0.0105)	0.8253(0.0120)	0.9106(0.0025)	0.9046(0.0022)	0.9516(0.0015)
D4	0.8417(0.0041)	0.8010(0.0108)	0.8201(0.0182)	0.9059(0.0011)	0.8948(0.0024)	0.9372(0.0016)
D5	0.7352(0.0055)	0.8270(0.0081)	0.8188(0.0147)	0.8570(0.0022)	0.7909(0.0038)	0.9069(0.0025)
D6	0.7757(0.0074)	0.8496(0.0169)	0.8346(0.0147)	0.8901(0.0028)	0.8326(0.0073)	0.9203(0.0004)
D7	0.8265(0.0046)	0.8518(0.0179)	0.8151(0.0073)	0.9043(0.0016)	0.9069(0.0022)	0.9334(0.0009)
D8	0.8532(0.0037)	0.8203(0.0178)	0.8459(0.0134)	0.9123(0.0025)	0.9091(0.0023)	0.9442(0.0007)
D9	0.8154(0.0071)	0.8452(0.0124)	0.8048(0.0069)	0.8927(0.0025)	0.8796(0.0020)	0.9253(0.0009)
Ave	0.8219	0.8303	0.8285	0.8997	0.8818	0.9338
Ave. std	0.0052	0.0138	0.0130	0.0020	0.0031	0.0012

TABLE 8: Results of all methods on the AUC index.

Datasets	ELM	W1-ELM	W2-ELM	R1-ELM	R2-ELM	G-ELM
D1	0.8542(0.0035)	0.8371(0.0088)	0.8679(0.0167)	0.9175(0.0009)	0.9125(0.0034)	0.9458(0.0012)
D2	0.8367(0.0042)	0.8418(0.0159)	0.8438(0.0117)	0.9117(0.0016)	0.9067(0.0023)	0.9408(0.0015)
D3	0.8750(0.0055)	0.8409(0.0102)	0.8319(0.0120)	0.9108(0.0025)	0.9050(0.0022)	0.9517(0.0015)
D4	0.8442(0.0042)	0.8162(0.0113)	0.8546(0.0176)	0.9067(0.0011)	0.8950(0.0024)	0.9375(0.0016)
D5	0.7372(0.0054)	0.8296(0.0070)	0.8239(0.0128)	0.8607(0.0020)	0.7939(0.0035)	0.9089(0.0025)
D6	0.7839(0.0071)	0.8616(0.0156)	0.8403(0.0147)	0.8922(0.0027)	0.8339(0.0072)	0.9217(0.0004)
D7	0.8322(0.0041)	0.8569(0.0166)	0.8165(0.0067)	0.9056(0.0016)	0.9072(0.0022)	0.9344(0.0009)
D8	0.8583(0.0037)	0.8319(0.0167)	0.8561(0.0114)	0.9128(0.0025)	0.9094(0.0023)	0.9444(0.0007)
D9	0.8275(0.0062)	0.8494(0.0119)	0.7990(0.0070)	0.8938(0.0026)	0.8794(0.0020)	0.9263(0.0010)
Ave	0.8277	0.8406	0.8371	0.9013	0.8826	0.9346
Ave. std	0.0049	0.0127	0.0123	0.0019	0.0031	0.0013

4.4. *Experimental Results.* To evaluate the classification performance of the proposed G-ELM, five ELM-based methods were used for performance comparison. All experiments were repeated ten times for fairness. The mean and standard deviation of the corresponding indices of all methods in each dataset are reported in Tables 3–8. The best results are shown in bold. The improvement of G-ELM relative to ELM on all

datasets using the six imbalanced classification evaluation indices is shown in Figure 5.

According to experimental results in Tables 3–8, the following observations can be made:

- (1) For the adopted six imbalanced classification evaluation indices, the proposed G-ELM performs best on

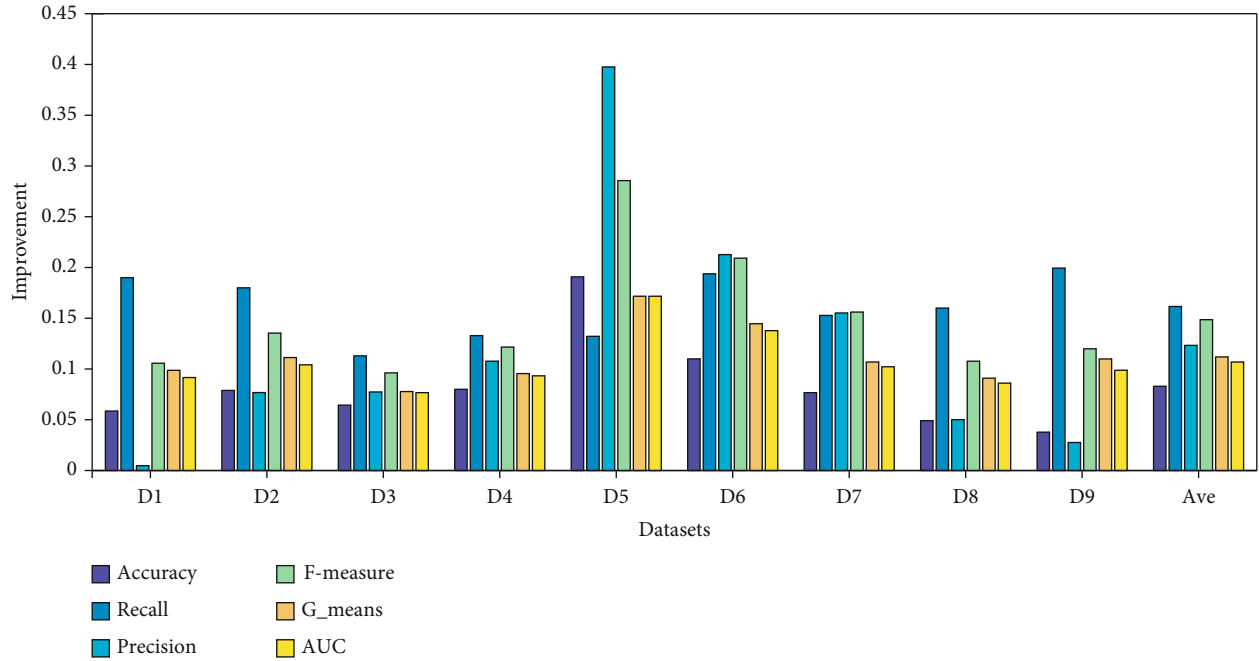


FIGURE 5: Improvement of G-ELM compared with ELM across six indices on all datasets.

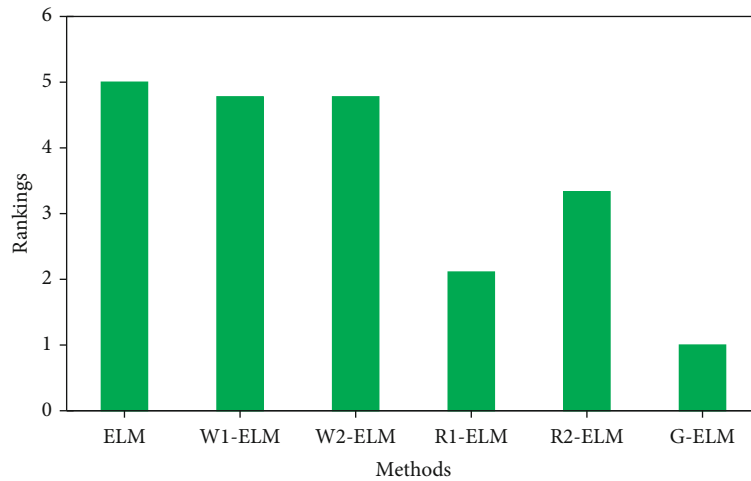


FIGURE 6: Rankings of all methods (Friedman test).

most datasets. This is because G-ELM can suppress the misclassification of negative samples while considering the accuracy of positive samples and has a high classification performance, which is suitable for imbalanced epileptic EEG signal recognition

- (2) In general, G-ELM, R1-ELM, R2-ELM, W1-ELM, and W2-ELM achieved better performances than ELM. This is due to the addition of a cost matrix, which makes them more suitable for imbalanced classification. Moreover, G-ELM has the best effect because it adds sample information using a relationship graph
- (3) Tables 4 and 5 show the results of recall and precision of all methods. They evaluate the performance of

TABLE 9: The post hoc hypothesis test results ( $\alpha_{\text{Fri}} = 0.05$ ).

$i$	Methods	$z$	$p_{\text{Fri}}$	Holm	Hypothesis
5	ELM	4.535574	0.000006	0.01	Rejected
4	W1-ELM	4.283597	0.000018	0.0125	Rejected
3	W2-ELM	4.283597	0.000018	0.016667	Rejected
2	R2-ELM	2.645751	0.008151	0.025	Rejected
1	R1-ELM	1.259882	0.207712	0.05	Not rejected

imbalanced classification from two different perspectives. From the excellent performance of G-ELM in Tables 4 and 5, we can see that adding information regarding relationships between samples can increase the weight of positive samples

- (4)  $F$ -measure and  $G$ \_means in Tables 6 and 7 are two important indices to measure the performance of imbalanced classification methods, which can be combined with recall and precision to evaluate the effect of the methods. From the results, we can see that the proposed G-ELM has the best performance. It has excellent performance in imbalanced epileptic EEG signal recognition
- (5) AUC is an important index to evaluate imbalanced classifiers. From Table 8, we can see that the performance of G-ELM on all datasets is the best. G-ELM has excellent performance in imbalanced classification and good effectiveness of imbalanced epileptic EEG signal recognition

**4.5. Statistical Analysis.** Statistical analysis was performed to further analyze the performances of all the adopted methods in our experiments. For conciseness, we only present statistical analysis of the  $G$ \_means results. Firstly, the Friedman test [47] was used to calculate the average ranking of each method. The rankings of all the adopted methods are shown in Figure 6. In Figure 6, we can see that the performance of G-ELM is the best.

Then, the post hoc hypothesis test [48] was used to evaluate the statistical significance of the performance differences between G-ELM and the other adopted methods. Post hoc hypothesis test results ( $\alpha_{\text{Fri}} = 0.05$ ) are presented in Table 9. In Table 9, we can see that the null hypothesis is rejected when  $p_{\text{Fri}} \leq 0.025$  due to  $p_{\text{Fri}} \leq \text{Holm}$ . Therefore, performance differences between G-ELM and the other adopted methods are significant, which means that G-ELM is effective for imbalanced epileptic EEG signal recognition.

## 5. Conclusions

In this study, we aimed to address the challenge that traditional machine learning methods ignore the imbalance of epileptic EEG datasets and the existing imbalanced classification methods ignore the relationships between samples. A graph-based ELM was proposed for imbalanced epileptic EEG signal recognition. First, graph theory was used to construct the relationship between samples according to the distribution. Second, a model combining the relationship graph and ELM was proposed; this model inherited the rapid learning and good generalization capabilities of ELM while maintaining satisfactory classification. Experiments on a real imbalanced epileptic EEG dataset demonstrated the effectiveness and applicability of the proposed method. However, there is still room for improvement in the scope and search method of the optimal parameters in this experiment. In the future, ways to design a better method to determine the optimal parameters will be further studied and explored.

## Data Availability

Data can be downloaded from <http://www.epileptologie-bonn.de/cms/upload/workgroup/lehnerz/eegdata.html>.

## Conflicts of Interest

None of the authors have any conflicts of interest.

## Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 61772198 and by the Natural Science Foundation of Jiangsu Province under Grant BK20161268.

## References

- [1] S. Benbadis and W. Hauser, "An estimate of the prevalence of psychogenic non-epileptic seizures," *Seizure*, vol. 9, no. 4, pp. 280–281, 2000.
- [2] C. Yang, Z. Deng, K. Choi, and S. Wang, "Takagi–Sugeno–Kang transfer learning fuzzy logic system for the adaptive recognition of epileptic electroencephalogram signals," *IEEE Transactions on Fuzzy Systems*, vol. 24, no. 5, pp. 1079–1094, 2016.
- [3] Z. Jiang, F.-L. Chung, and S. Wang, "Recognition of multi-class epileptic EEG signals based on knowledge and label space inductive transfer," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 4, pp. 630–642, 2019.
- [4] M. Z. Parvez and M. Paul, "Epileptic seizure prediction by exploiting spatiotemporal relationship of EEG signals using phase correlation," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 24, no. 1, pp. 158–168, 2016.
- [5] Z. Iscan, Z. Dokur, and T. Demiralp, "Classification of electroencephalogram signals with combined time and frequency features," *Expert Systems with Applications*, vol. 38, no. 8, pp. 10499–10505, 2011.
- [6] T. Tzallas, M. G. Tsipouras, and D. I. Fotiadis, "Epileptic seizure detection in eegs using time-frequency analysis," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 5, pp. 703–710, 2009.
- [7] Z. Yang, K. C. Deng, K.-S. Choi, Y. Jiang, and S. Wang, "Transductive domain adaptive learning for epileptic electroencephalogram recognition," *Artificial Intelligence in Medicine*, vol. 62, no. 3, pp. 165–177, 2014.
- [8] Y. Zhang, J. Dong, J. Zhu, and C. Wu, "Common and special knowledge-driven TSK fuzzy system and its modeling and application for epileptic EEG signals recognition," *IEEE Access*, vol. 7, pp. 127600–127614, 2019.
- [9] F. P. Lestari, M. Haekal, R. E. Edison, F. R. Fauzy, S. N. Khotimah, and F. Haryanto, "Epileptic Seizure Detection in EEGs by Using Random Tree Forest, Naïve Bayes and KNN Classification," *Journal of Physics: Conference Series*, vol. 1505, no. 1, p. 12055, 2020.
- [10] H. Choubey and A. Pandey, "A combination of statistical parameters for the detection of epilepsy and EEG classification using ANN and KNN classifier," *Signal, Image and Video Processing*, vol. 15, no. 3, pp. 475–483, 2021.
- [11] Y. Zhang, Y. Zhang, J. Wang, and X. Zheng, "Comparison of classification methods on EEG signals based on wavelet packet decomposition," *Neural Computing and Applications*, vol. 26, no. 5, pp. 1217–1225, 2015.
- [12] Y. Zhang, H. Ishibuchi, and S. Wang, "Deep Takagi–Sugeno–Kang fuzzy classifier with shared linguistic fuzzy rules," *IEEE*

- Transactions on Fuzzy Systems*, vol. 26, no. 3, pp. 1535–1549, 2018.
- [13] Z. Deng, P. Xu, L. Xie, K.-S. Choi, and S. Wang, “Transductive joint-knowledge-transfer TSK FS for recognition of epileptic EEG signals,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 8, pp. 1481–1494, 2018.
- [14] Q. Yuan, W. Zhou, S. Li, and D. Cai, “Epileptic EEG classification based on extreme learning machine and nonlinear features,” *Epilepsy Research*, vol. 96, no. 1-2, pp. 29–38, 2011.
- [15] L. Shi and B. Lu, “EEG-based vigilance estimation using extreme learning machines,” *Neurocomputing*, vol. 102, no. 15, pp. 135–143, 2013.
- [16] Q. Yuan, W. Zhou, L. Zhang et al., “Epileptic seizure detection based on imbalanced classification and wavelet packet transform,” *Seizure*, vol. 50, pp. 99–108, 2017.
- [17] M. K. Siddiqui, X. Huang, R. Morales-Menendez, N. Hussain, and K. Khatoun, “Machine learning based novel cost-sensitive seizure detection classifier for imbalanced EEG data sets,” *International Journal on Interactive Design and Manufacturing*, vol. 14, no. 4, pp. 1491–1509, 2020.
- [18] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, “Extreme learning machine: a new learning scheme of feedforward neural networks,” in *2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No.04CH37541)*, vol. 2, pp. 985–990, Budapest, Hungary, 2004.
- [19] G.-B. Huang, Q. Y. Zhu, and C. K. Siew, “Extreme learning machine: theory and applications,” *Neurocomputing*, vol. 70, no. 1-3, pp. 489–501, 2006.
- [20] W. Zong, G.-B. Huang, and Y. Chen, “Weighted extreme learning machine for imbalance learning,” *Neurocomputing*, vol. 101, pp. 229–242, 2013.
- [21] W. B. Zhang and H. B. Ji, “Fuzzy extreme learning machine for classification,” *Electronics Letters*, vol. 49, no. 7, pp. 448–450, 2013.
- [22] H. Yu, C. Sun, X. Yang, W. Yang, J. Shen, and Y. Qi, “ODOC-ELM: optimal decision outputs compensation-based extreme learning machine for classifying imbalanced data,” *Knowledge-Based Systems*, vol. 92, pp. 55–70, 2016.
- [23] K. Li, X. Kong, Z. Lu, L. Wenyin, and J. Yin, “Boosting weighted elm for imbalanced learning,” *Neurocomputing*, vol. 128, pp. 15–21, 2014.
- [24] J. Yang, H. Yu, X. Yang, and X. Zuo, “Imbalanced extreme learning machine based on probability density estimation,” in *International Workshop on Multi-disciplinary Trends in Artificial Intelligence*, pp. 160–167, Springer, Cham, 2015.
- [25] S. Shukla and R. N. Yadav, “Regularized weighted circular complex-valued extreme learning machine for imbalanced learning,” *IEEE Access*, vol. 3, pp. 3048–3057, 2015.
- [26] W. Xiao, J. Zhang, Y. Li, and W. Yang, “Imbalanced extreme learning machine for classification with imbalanced data distributions,” *Proceedings in Adaptation, Learning and Optimization*, vol. 7, pp. 503–514, 2016.
- [27] J. Du, C. Vong, Y. Chang, and Y. Jiao, “Online sequential extreme learning machine with under-sampling and over-sampling for imbalanced big data classification,” *Proceedings of ELM-2016*, 2018.
- [28] Y. Zhang, B. Liu, J. Cai, and S. Zhang, “Ensemble weighted extreme learning machine for imbalanced data classification based on differential evolution,” *Neural Computing and Applications*, vol. 28, Supplement 1, pp. 259–267, 2017.
- [29] W. Xiao, J. Zhang, Y. Li, S. Zhang, and W. Yang, “Class-specific cost regulation extreme learning machine for imbalanced classification,” *Neurocomputing*, vol. 261, pp. 70–82, 2017.
- [30] H. Yu, C. Sun, X. Yang, S. Zheng, Q. Wang, and X. Xi, “LW-ELM: a fast and flexible cost-sensitive learning framework for classifying imbalanced data,” *IEEE Access*, vol. 6, pp. 28488–28500, 2018.
- [31] B. S. Raghuwanshi and S. Shukla, “Class-specific extreme learning machine for handling binary class imbalance problem,” *Neural Networks*, vol. 105, pp. 206–217, 2018.
- [32] R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. E. Elger, “Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state,” *Physical Review E*, vol. 64, no. 6, article 061907, 2001.
- [33] T. Strutz, *Data Fitting and Uncertainty: A Practical Introduction to Weighted Least Squares and Beyond*, Vieweg, Wiesbaden, Germany, 2010.
- [34] G. Li, K. Chang, and S. C. H. Hoi, “Multiview semi-supervised learning with consensus,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 11, pp. 2040–2051, 2012.
- [35] F. Zhu, L. Shao, and M. Lin, “Multi-view action recognition using local similarity random forests and sensor fusion,” *Pattern Recognition Letters*, vol. 34, no. 1, pp. 20–24, 2013.
- [36] B. R. Greene, S. Faul, W. P. Marnane, G. Lightbody, I. Korotchikova, and G. B. Boylan, “A comparison of quantitative EEG features for neonatal seizure detection,” *Clinical Neurophysiology*, vol. 119, no. 6, pp. 1248–1261, 2008.
- [37] E. A. Vivaldi and A. Bassi, “Frequency domain analysis of sleep eeg for visualization and automated state detection,” in *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 3740–3743, New York, NY, USA, August 2006.
- [38] C. W. N. F. C. W. Fadzal, W. Mansor, L. Y. Khuan, and A. Zabidi, “Short-time Fourier transform analysis of EEG signal from writing,” in *2012 IEEE 8th International Colloquium on Signal Processing and its Applications*, pp. 525–527, Malacca, Malaysia, March 2012.
- [39] H. Ocak, “Automatic detection of epileptic seizures in EEG using discrete wavelet transform and approximate entropy,” *Expert Systems with Applications*, vol. 36, no. 2, pp. 2027–2036, 2009.
- [40] D. Hu, W. Li, and X. Chen, “Feature extraction of motor imagery EEG signals based on wavelet packet decomposition,” in *The 2011 IEEE/ICME International Conference on Complex Medical Engineering*, pp. 694–697, Harbin, China, May 2011.
- [41] M. Leshno, V. Y. Lin, A. Pinkus, and S. Schocken, “Multilayer feedforward networks with a nonpolynomial activation function can approximate any function,” *Neural Networks*, vol. 6, no. 6, pp. 861–867, 1993.
- [42] Z. Zhu, Z. Wang, D. Li, and W. Du, “Multiple empirical kernel learning with majority projection for imbalanced problems,” *Applied Soft Computing*, vol. 76, pp. 221–236, 2019.
- [43] O. M. Younis, M. M. Krunz, and S. Ramasubramanian, “ROC: resilient online coverage for surveillance applications,” *IEEE/ACM Transactions on Networking*, vol. 19, no. 1, pp. 251–264, 2011.

- [44] Y. Zhang, S. Wang, K. Xia, Y. Jiang, and P. Qian, "Alzheimer's disease multiclass diagnosis via multimodal neuroimaging embedding feature selection and fusion," *Information Fusion*, vol. 66, pp. 170–183, 2021.
- [45] T. A. Joachims, "A support vector method for multivariate performance measures," in *Proceedings of the 22nd International Conference on Machine Learning - ICML '05*, pp. 377–384, Bonn, Germany, 2005.
- [46] J. Zhou, Z. Jiang, and S. Wang, "Laplacian least learning machine with dynamic updating for imbalanced classification," *Applied Soft Computing*, vol. 88, p. 106028, 2020.
- [47] Y. Zhang, F.-L. Chung, and S. Wang, "Takagi-sugeno-kang fuzzy systems with dynamic rule weights," *Journal of Intelligent Fuzzy Systems*, vol. 37, no. 6, pp. 8535–8550, 2019.
- [48] Z. Deng, K.-S. Choi, Y. Jiang, and S. Wang, "Generalized hidden-mapping ridge regression, knowledge-leveraged inductive transfer learning for neural networks, fuzzy systems and kernel methods," *IEEE Transactions on Cybernetics*, vol. 44, no. 12, pp. 2585–2599, 2014.

## Research Article

# Evaluation and Prediction of COVID-19 Prevention and Control Strategy Based on the SEIR-AQ Infectious Disease Model

Yue Yu , Yuxing Zhou, Xiangzhong Meng, Wenfei Li, Yang Xu, Manfeng Hu, and Jingxiang Zhang 

School of Science, Jiangnan University, 1800 Lihu Avenue, Wuxi, Jiangsu 214122, China

Correspondence should be addressed to Jingxiang Zhang; zhangjingxiang@jiangnan.edu.cn

Received 8 May 2021; Revised 23 May 2021; Accepted 31 May 2021; Published 11 June 2021

Academic Editor: Shan Zhong

Copyright © 2021 Yue Yu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Based on the SEIR model, which takes into account prevention and control measures, prevention and control awareness, and economic level and medical level indicators, this paper proposes an infectious disease model of “susceptible-exposed-infected-removed-asymptomatic-isolated” (short for SEIR-AQ) to assess and predict the development of the COVID-19 pandemic with different prevention and control strategies. The kinetic parameters of the SEIR-AQ model were obtained by fitting, and the parameters of the SEIR-AQ model were solved through the Euler method. Furthermore, the effects of different countries’ prevention and control strategies on the number of infections, the proportion of isolation, the number of deaths, and the number of recoveries were also simulated. The theoretical analysis showed that measures such as isolation for prevention and control and medical tracking isolation had a significant inhibitory effect on the development of the COVID-19 pandemic, among which stratified treatment and enhanced awareness played a key role in the rapid regression of the peak of COVID-19-infected patients. *Conclusion of the Simulation.* The SEIR-AQ model can be used to evaluate the development status of the COVID-19 epidemic and has some theoretical value for the prediction of COVID-19.

## 1. Introduction

Coronavirus disease 2019 is a lung disease caused by SARS-CoV-2, which is highly lethal and infectious [1]. Some scholars have clinically analyzed the causes of death in patients with confirmed COVID-19 [2] and identified factors associated with the death of patients with COVID-19 pneumonia caused by the novel coronavirus SARS-CoV-2 [3]. With 130566186 cumulative diagnoses and 2842363 cumulative deaths worldwide as of April 4, 2021, COVID-19 not only impacted the global economy but also deeply affected the governance of all countries in the world.

For COVID-19 epidemic prediction, most scholars used the classical SIR and SEIR models proposed by Beretta and Takeuchi [4] and Cooke and van den Driessche [5] to infer the COVID-19 peak time and maximum number of confirmed cases based on the existing data. Yu et al. [6] evaluated and predicted COVID-19 based on a SIR model with time-varying parameters to obtain expected inflection points and

maximum number of confirmed cases. Wei et al. [7] and Geng et al. [8] studied the effect of prevention and control isolation measures on the development trend of the COVID-19 epidemic based on the SEIR model and concluded that strict prevention and control isolation measures can slow down the development trend of the COVID-19 epidemic. Wang et al. [9] established the SEIADR model by introducing asymptomatic infected individuals on the traditional SEIR model. Also, they predicted the development of the COVID-19 epidemic in Hubei Province, which had better fitting effect compared with the SEIR model. Shao et al. [10] used the classical SEIR model to conduct a predictive analysis of COVID-19 in Shandong Province and Korea, comparing the impact of control measures on the spread of COVID-19. Li et al. [11] fitted the COVID-19 regeneration coefficient ( $R_0$ ) curve based on the SEIR model to predict and analyze the development trend of the COVID-19 epidemic in Hubei Province, China, America, India, Italy, and Iran and also predicted that the spread of the epidemic in Hubei Province

would be better controlled compared with that of foreign countries. Lin [12], Chen et al. [13], and Ansumali et al. [14] introduced asymptomatic infected individuals based on the SEIR model, which led to a significant improvement in the fitting and prediction performance of the SEIR model. Ivorra et al. [15] developed the  $\theta$ -SEIHRD model, where asymptomatic infected patients and medical conditions were taken into account, which could predict hospital bed demand more accurately. Pai et al. [16] improved the SEIR model by integrating government control policies, public health, and other factors to analyze the impact on the development trend of COVID-19 in India.

Although the above-mentioned studies have achieved certain effects, they only consider the latent infectious capacity of COVID-19-infected people, the infectious capacity of asymptomatic patients, prevention and quarantine measures, etc. In fact, many other factors including different medical levels, economic levels, and prevention and control awareness also have a great impact on the transmission of COVID-19. Therefore, based on the SEIR model, we divided susceptible people, contacts, and infected people into isolated and exposed states and then introduced hospitalized patients and asymptomatic patients as well as indicators such as prevention and control measures, prevention and control awareness, economic level, and medical level to construct a more interpretive SEIR-AQ paradigm. The development trend of COVID-19 at different levels was simulated by adjusting the parameters of prevention and control measures, prevention and control awareness, economic level, and medical level.

## 2. Establishment of the SEIR-AQ Model

The traditional SEIR model divides the population into susceptible people ( $S$ ), contacts ( $E$ ), infected people ( $I$ ), and recovered people ( $R$ ). However, the SEIR-AQ paradigm adds isolation of susceptible people ( $S_q$ ), isolation of contacts ( $E_q$ ), isolation of infected people ( $I_q$ ), asymptomatic patients ( $A$ ), and hospitalized patients ( $H$ ). Now, according to the proportion of  $b_1$ ,  $b_2$ , and  $b_3$ , we converted the isolated infected people, unisolated infected people, and asymptomatic patients into hospitalized patients  $H$ . The SEIR-AQ paradigm assumes that the infected and quarantined people are not infectious during the isolation period, and the infected people are immune to cure. The parameter  $\nu$  is the ratio of the transmission capacity of the unisolated contacts  $E$  to the unisolated infected people  $I$ , and the parameter  $\theta$  is the ratio of the transmission capacity of the asymptomatic patients  $A$  to the unisolated infected people  $I$ . Therefore, the SEIR-AQ paradigm in this paper is more interpretive and adaptable. The warehouse conversion relationship of the SEIR-AQ paradigm is shown in Figure 1.

$q$ ,  $\beta$ ,  $c$ , and  $\rho$  are the isolation ratio, the infection probability, the contact rate, and the effective contact coefficient, respectively, and  $\rho c$  is the effective contact rate. The conversion rate from unisolated susceptible people ( $S$ ) to isolation of  $S_q$ ,  $E_q$ , and unisolated contacts ( $E$ ) is  $\rho c q(1 - \beta)$ ,  $\rho c q \beta$ , and  $\rho c \beta(1 - q)$ , respectively. At the same time, considering

the impact of unisolated infected people ( $I$ ),  $A$ , and  $E$  on susceptible populations, there is also isolation of  $S_q$  that is retransformed into  $S$  at a rate of  $\lambda$ . The natural death rate of  $S$  is  $\eta$ . Therefore, the governing equation for the number of susceptible people is

$$\frac{dS}{dt} = -[\rho c \beta + \rho c q(1 - \beta)]S(I + \theta A + \nu E) + \lambda S_q - \eta S. \quad (1)$$

$\lambda$  is the quarantine release rate, taking  $\lambda = 1/14$  (the quarantine duration is 14 days).

The SEIR-AQ model considers different prevention and control measures, prevention and control awareness, economic level, and medical level:

$$\left\{ \begin{array}{l} \frac{dS}{dt} = -[\rho c \beta + \rho c q(1 - \beta)]S(I + \theta A + \nu E) + \lambda S_q - \eta S, \\ \frac{dE}{dt} = \rho c \beta(1 - q)S(I + \theta A + \nu E) - (\sigma_1 + \sigma_2 + \sigma_3 + \eta)E, \\ \frac{dI}{dt} = -(b_1 + r_1 + \alpha_1)I + \sigma_1 e E + \delta_1 e E_q + p_1 A, \\ \frac{dA}{dt} = -(p_1 + p_2 + b_2 + r_2 + \alpha_2)A + \sigma_2(1 - e)E + \delta_2(1 - e)E_q, \\ \frac{dS_q}{dt} = \rho c q(1 - \beta)S(I + \theta A + \nu E) - (\lambda + \eta)S_q, \\ \frac{dE_q}{dt} = \rho c \beta q S(I + \theta A + \nu E) - (\delta_1 + \delta_2 + \delta_3 + \eta)E_q, \\ \frac{dI_q}{dt} = -(b_3 + r_3 + \alpha_3)I_q + \sigma_3 e E + \delta_3 e E_q + p_2 A, \\ \frac{dH}{dt} = b_1 I + b_2 A + b_3 I_q - (r_4 + \alpha_4)H, \\ \frac{dR}{dt} = \mu_1 r_1 I + \mu_2 r_2 A + \mu_3 r_3 I_q + \mu_4 r_4 H - \eta R. \end{array} \right. \quad (2)$$

$\sigma_1$ ,  $\sigma_2$ , and  $\sigma_3$  are the rates of conversion of  $E$  to  $I$ ,  $A$ , and  $I_q$ , respectively, taking  $\sigma_1 = \sigma_2 = \sigma_3 = 7$  (the incubation period is 7 days);  $\alpha_1$ ,  $\alpha_2$ ,  $\alpha_3$ , and  $\alpha_4$  are the death rates of  $I$ ,  $A$ ,  $I_q$ , and  $H$ ;  $\delta_1$ ,  $\delta_2$ , and  $\delta_3$  are the rates of isolation of  $E_q$  to  $I$ ,  $A$ , and  $I_q$ , respectively;  $p_1$  and  $p_2$  are the rates at which  $A$  turns into  $I$  and  $I_q$ ;  $r_1$ ,  $r_2$ ,  $r_3$ , and  $r_4$  are the recovery rate of  $I$ ,  $A$ ,  $I_q$ , and  $H$ ;  $\mu_1$ ,  $\mu_2$ ,  $\mu_3$ , and  $\mu_4$  are the coefficient of the recovery rate of  $I$ ,  $A$ ,  $I_q$ , and  $H$ ; and  $e$  is the probability that the infected people have symptoms.

In fact,  $I_q$  will be immediately sent to a designated hospital for isolation and treatment during the COVID-19 epidemic; thus,  $I_q$  will all be converted into  $H$  in this model based on the SEIR-AQ model paradigm. Considering that the asymptomatic patients would not be taken to the hospital until any symptom was shown, we removed the asymptomatic patients into the relationship between the hospitalized patients and the infected people.  $E_q$ , if confirmed, will be directly sent to the hospital and converted into  $H$ . Therefore, the relationship that  $E_q$  is converted to  $I$  is removed. Based on

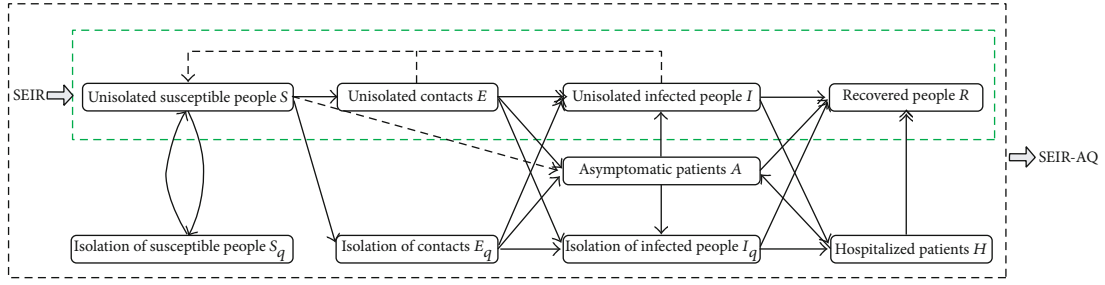


FIGURE 1: The warehouse conversion relationship of the SEIR-AQ paradigm.

the modified SEIR model established by Cao et al. [17] comprehensively considering the transmission characteristics of COVID-19, the population transformation relationship of SEIR-AQ, which is suitable for analyzing the COVID-19 epidemic situation, is shown in Figure 2. In Figure 2, the gray lines represent the transformation relationship deleted from the original and the red line represents the increased part converted from  $E_q$  to  $H$ .

At this time, the SEIR-AQ model is

$$\begin{cases} \frac{dS}{dt} = -[c\beta + cq(1 - \beta)]S(I + \theta A + vE) + \lambda S_q - \eta S, \\ \frac{dE}{dt} = c\beta(1 - q)S(I + \theta A + vE) - (\sigma + \eta)E, \\ \frac{dI}{dt} = -(b + r_1 + \alpha + \eta)I + \sigma eE, \\ \frac{dA}{dt} = -(r_2 + \eta)A + \sigma(1 - e)E, \\ \frac{dS_q}{dt} = cq(1 - \beta)S(I + \theta A + vE) - (\lambda + \eta)S_q, \\ \frac{dE_q}{dt} = c\beta qS(I + \theta A + vE) - (\delta + \eta)E_q, \\ \frac{dH}{dt} = bI + \delta E_q - (r_4 + \alpha + \eta)H, \\ \frac{dR}{dt} = r_1 I + r_2 A + r_4 H - \eta R. \end{cases} \quad (3)$$

The parameter  $q$  represents the isolation ratio,  $\beta$  represents the probability of infection,  $c$  represents the contact

rate,  $\alpha$  represents the rate of death due to illness,  $\delta$  represents the rate where  $E_q$  is converted to  $H$ ,  $r$  represents the rate of recovery,  $e$  represents the probability that the infected people have symptoms,  $\lambda$  represents the rate of  $S_q$  converted to  $S$ ,  $\theta$  represents the ratio of the transmission capacity of  $A$  compared with  $I$ , and  $\eta$  represents the natural mortality rate.

### 3. Global Stability of the Equilibrium Point

In the proof of this paper, the following equation is considered apart from  $R(t)$  [18]:

$$\begin{cases} \frac{dS}{dt} = -[c\beta + cq(1 - \beta)]S(I + \theta A + vE) + \lambda S_q - \eta S, \\ \frac{dE}{dt} = c\beta(1 - q)S(I + \theta A + vE) - (\sigma + \eta)E, \\ \frac{dI}{dt} = -(b + r_1 + \alpha + \eta)I + \sigma eE, \\ \frac{dA}{dt} = -(r_2 + \eta)A + \sigma(1 - e)E, \\ \frac{dS_q}{dt} = cq(1 - \beta)S(I + \theta A + vE) - (\lambda + \eta)S_q, \\ \frac{dE_q}{dt} = c\beta qS(I + \theta A + vE) - (\delta + \eta)E_q, \\ \frac{dH}{dt} = bI + \delta E_q - (r_4 + \alpha + \eta)H. \end{cases} \quad (4)$$

According to model (4), we can express  $f(I)$  as

$$f(I) = I - \frac{e\sigma\eta(\lambda + \eta)(r_2 + \eta)}{c[\lambda\beta + \eta(\beta + q - \beta q)]\{\sigma e(r_2 + \eta) + [\theta\sigma(1 - e) + v(r_2 + \eta)](b + r_1 + \alpha + \eta)\}}. \quad (5)$$

Then, the derivation with respect to  $I$  is computed as

$$f'(I) = 1 > 0. \quad (6)$$

It shows that  $f(I)$  is an increasing function as the density of  $I$  tends to go infinity, which implies that

$$\lim_{I \rightarrow +\infty} f(I) = +\infty. \quad (7)$$



According to equation (5), we can obtain

$$f(0) = -\frac{e\sigma\eta(\lambda + \eta)(r_2 + \eta)}{c[\lambda\beta + \eta(\beta + q - \beta q)]\{\sigma e(r_2 + \eta) + [\theta\sigma(1 - e) + \nu(r_2 + \eta)](b + r_1 + \alpha + \eta)\}} < 0. \quad (8)$$

So the equilibrium point exists.

$$F = \begin{bmatrix} 0 & 0 & 0 & 0 & \lambda & 0 & 0 & 0 \\ c\beta(1 - q)(I + \theta A + \nu E) & c\beta(1 - q)\nu S & c\beta(1 - q)S & c\beta(1 - q)\theta S & 0 & 0 & 0 & 0 \\ 0 & \sigma e & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma(1 - e) & 0 & 0 & 0 & 0 & 0 & 0 \\ cq(1 - \beta)(I + \theta A + \nu E) & cq(1 - \beta)\nu S & cq(1 - \beta)S & cq(1 - \beta)\theta S & 0 & 0 & 0 & 0 \\ cq\beta(I + \theta A + \nu E) & cq\beta\nu S & cq\beta S & cq\beta\theta S & 0 & 0 & 0 & 0 \\ 0 & 0 & b & 0 & 0 & \delta & 0 & 0 \end{bmatrix}, \quad (9)$$

$$V = \begin{bmatrix} [c\beta + cq(1 - \beta)](I + \theta A + \nu E) + \eta & [c\beta + cq(1 - \beta)]\nu S & [c\beta + cq(1 - \beta)]S & [c\beta + cq(1 - \beta)]\theta S & 0 & 0 & 0 & 0 \\ 0 & \sigma + \eta & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & b = r_1 + \alpha + \eta & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & r_2 + \eta & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda + \eta & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \delta + \eta & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & r_4 + \alpha + \eta \end{bmatrix}, \quad (10)$$

$$R_0 = \rho(FV^{-1}). \quad (11)$$

$R_0$  is the basic reproduction number of model (4). Then, there comes Proposition 1.

**Proposition 1.** *Model (4) admits a disease-free equilibrium  $P^0 = (0, 0, 0, 0, 0, 0, 0)$ . If  $R_0 > 1$ , then model (4) admits a unique endemic equilibrium  $P^* = (S^*, E^*, I^*, A^*, S_q^*, E_q^*, H^*)$ :*

$$S^* = \frac{(\sigma + \eta)(r_2 + \eta)(b + r_1 + \alpha + \eta)}{c\beta(1 - q)\{\sigma e(r_2 + \eta) + [\theta\sigma(1 - e) + \nu(r_2 + \eta)](b + r_1 + \alpha + \eta)\}}, \quad (12)$$

$$E^* = -\frac{\eta(\lambda + \eta)(r_2 + \eta)(b + r_1 + \alpha + \eta)}{c[\lambda\beta + \eta(\beta + q - \beta q)]\{\sigma e(r_2 + \eta) + [\theta\sigma(1 - e) + \nu(r_2 + \eta)](b + r_1 + \alpha + \eta)\}}, \quad (13)$$

$$I^* = \frac{\sigma\eta(\lambda + \eta)(r_2 + \eta)}{c[\lambda\beta + \eta(\beta + q - \beta q)]\{\sigma e(r_2 + \eta) + [\theta\sigma(1 - e) + \nu(r_2 + \eta)](b + r_1 + \alpha + \eta)\}}, \quad (14)$$

$$A^* = \frac{\sigma\eta(1-e)(\lambda+\eta)(b+r_1+\alpha+\eta)}{c[\lambda\beta+\eta(\beta+q-\beta q)]\{\sigma e(r_2+\eta)+[\theta\sigma(1-e)+\nu(r_2+\eta)](b+r_1+\alpha+\eta)\}}, \quad (15)$$

$$S_q^* = \frac{q\eta(1-\beta)(\sigma+\eta)(r_2+\eta)(b+r_1+\alpha+\eta)}{c\beta(1-q)[\lambda\beta+\eta(\beta+q-\beta q)]\{\sigma e(r_2+\eta)+[\theta\sigma(1-e)+\nu(r_2+\eta)](b+r_1+\alpha+\eta)\}}, \quad (16)$$

$$H^* = \frac{\eta(\lambda+\eta)(r_2+\eta)[b\sigma e(1-q)(\delta+\eta)+\delta q(\sigma+\eta)(b+r_1+\alpha+\eta)]}{c(r_4+\alpha+\eta)(\delta+\eta)(1-q)[\lambda\beta+\eta(\beta+q-\beta q)]\{\sigma e(r_2+\eta)+[\theta\sigma(1-e)+\nu(r_2+\eta)](b+r_1+\alpha+\eta)\}}. \quad (17)$$

Next, the global stabilities of the equilibria of model (4) around the disease-free equilibrium and the endemic equilibrium are investigated in Theorem 2, respectively.

**Theorem 2.** *If  $R_0 \leq 1$ , then the disease-free equilibrium  $P^0 = (0, 0, 0, 0, 0, 0)$  of model (4) is globally asymptotically stable. If  $R_0 > 1$ , then the unique endemic equilibrium  $P^* = (S^*, E^*, I^*, A^*, S_q^*, E_q^*, H^*)$  of model (4) is globally asymptotically stable.*

#### 4. Model Parameter Assignment and Verification

To further illustrate the applicability of the SEIR-AQ model, we select COVID-19 data from China, America, Brazil, and India for analysis. The numbers of BRIC, China, Brazil, and India share the similar economic level while America boosts higher level. At the same time, China and India have relatively higher population density. America and Brazil have similar populations and prevention and control awareness. Based on the SEIR-AQ model and the values of different parameters, the impact of different prevention and control measures, prevention and control awareness, economic level, and medical level on the development trend of the COVID-19 epidemic is simulated.

*4.1. Parameter Setting and Fitting.* The initial values of the SEIR-AQ model were referenced to the single-day confirmed data of COVID-19 in four countries from January 23, 2020, to November 10, 2020. The parameters were estimated in conjunction with relevant literature. The initial values are shown in Table 1, and the parameter values are shown in Table 2. The evaluation index of the SEIR-AQ model fit was the coefficient of determination [12], which is shown in Table 3.

From Figure 3, it can be concluded that the number of single-day confirmed cases of the COVID-19 outbreak in China increased rapidly at the beginning of the outbreak, peaked in February 2020, and then decreased rapidly to a stable and manageable state. The COVID-19 outbreak in America started in April and soon reached to top on April 5 and July 20, respectively. The overall trend of single-day confirmed cases continued to increase. Meanwhile, when it comes to Brazil and India, the number continued to decrease after the peak at August 1 and September 26, respectively, and continued to decrease thereafter. The predictions based

on the SEIR-AQ model fit basically matched the actual epidemic development trend, and the SEIR-AQ model fitting evaluation index determination coefficient  $R^2$  is greater than 85%, indicating that the SEIR-AQ model has a significant fitting effect and can predict the development trend of the COVID-19 epidemic.

#### 5. Further Discussion of the SEIR-AQ Model

In this paper, we analyze and study COVID-19 based on the SEIR-AQ model, theoretically analyze the law of COVID-19 evolution, and analyze the influence of different countries' prevention and control measures, prevention and control awareness, economic level, and medical level on their COVID-19 evolution, in which the peak number of new confirmed cases in each country when the parameters change is shown in Table 4.

*5.1. Assessment of the Impact of Prevention and Control Measures and Prevention and Control Awareness on the COVID-19.* Since the emergence of the COVID-19 epidemic, the Chinese government has decisively adopted strict prevention and control measures to reduce  $c$  and increase  $q$ . Under the positive instructions of the Chinese government, the citizens gradually raise their awareness of epidemic prevention and control, which reduced  $\rho$  and minimized the development of clustered epidemics. As a result, the COVID-19 epidemic tends to be in a more controllable and stable state. In the initial period of the COVID-19 outbreak, only few effective prevention and control measures were taken by the United States government. This then lead to large  $c$  and small  $q$  which finally resulted in a significant increase in the number of the COVID-19 confirmed cases in a single day. Relatively weak awareness of prevention and control among U.S. nationals led to large  $\rho$ . Although America strengthened its prevention and control measures in the later period, it had missed the prevention and control window period, which led to the development of the COVID-19 epidemic. In April 2020, the COVID-19 epidemic developed on a large scale in Brazil. The lack of timely adoption of prevention and control measures proposed by WHO indirectly led to lower awareness of prevention and control among the Brazilian population, which increased  $c$  and  $\rho$  and lowered  $q$ . The number of single-day confirmed cases in Brazil continued to hit a record high. But in June, the government introduced compulsory epidemic prevention measures, such as the adoption of entry restrictions. The number of single-day confirmed

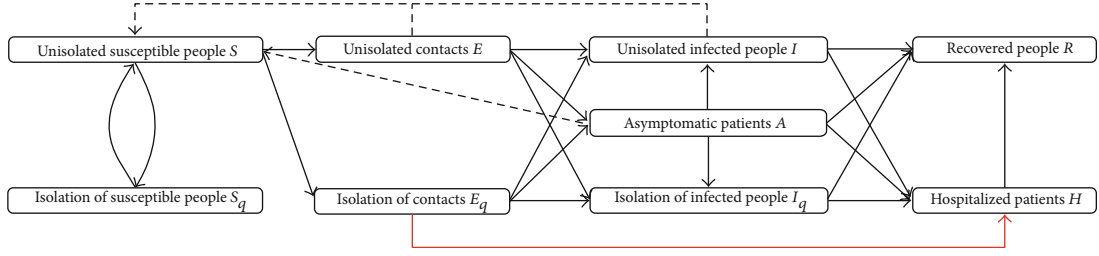


FIGURE 2: The warehouse conversion relationship of the SEIR-AQ model.

TABLE 1: SEIR-AQ model initial value setting.

Status	China	India	America	Brazil	Description
$S$	$14 \times 10^8$	$13.6 \times 10^8$	$3.28 \times 10^8$	$2.11 \times 10^8$	Total population data
$E$	4776	4314	5470	3904	Number of single-day confirmed cases
$I$	524	694	752	647	Official data
$R$	31	10	15	6	Official data
$S_q$	2776	1457	2652	1568	Official data, still under medical observation
$H$	924	854	1213	797	Patient isolation and partial medical observation, $I + E_q$
$E_q$	400	160	460	150	Estimated value, less than the number of people still under medical observation
$A$	262	347	376	324	Assuming that the undetected ratio is 0.5, $0.5I$

TABLE 2: SEIR-AQ model parameter values.

Parameter	China	India	America	Brazil	Description
$q$	$4.98 \times 10^{-10}$	$3.2 \times 10^{-9}$	$3.04 \times 10^{-9}$	$5.6 \times 10^{-9}$	Fitting optimization based on actual data
$\beta$	$1 \times 10^{-8}$	$0.7 \times 10^{-8}$	$2.5 \times 10^{-8}$	$0.5 \times 10^{-7}$	Fitting optimization based on actual data
$c$	2.2	3.8	3.5	3	Fitting optimization based on actual data
$\sigma$	1/7	1/7	1/7	1/7	The incubation period is set to 7 days
$\alpha$	$2.7 \times 10^{-4}$	$4.6 \times 10^{-2}$	$4.5 \times 10^{-4}$	$3.4 \times 10^{-2}$	Adjusting based on the actual number of deaths
$\delta$	0.13	0.017	0.076	0.03	Fitting optimization based on actual data
$r_1 = r_2 = r_4 = r$	$3.5 \times 10^{-2}$	$2 \times 10^{-4}$	$3 \times 10^{-6}$	$1.2 \times 10^{-3}$	Adjusting based on the actual number of people recovered
$e$	0.4	0.37	0.4	0.35	Fitting optimization based on actual data
$\lambda$	1/14	1/14	1/14	1/14	The quarantine period is set to 14 days
$\theta$	1	1	1	1	Assuming that the contact is the same as the patient who has shown symptoms
$\eta$	0	0	0	0	Assuming that the natural mortality rate is 0

TABLE 3: Analysis of the fitting degree of the SEIR-AQ model.

	$\bar{y}$	$SS_{tot}$	$SS_{res}$	$R^2$
China	$2.3482 \times 10^3$	$8.6654 \times 10^9$	$4.9524 \times 10^7$	0.9943
America	$3.5596 \times 10^4$	$1.9986 \times 10^8$	$2.0080 \times 10^7$	0.8995
Brazil	$2.1719 \times 10^4$	$8.0718 \times 10^7$	$6.3786 \times 10^6$	0.9210
India	$3.0041 \times 10^4$	$2.8920 \times 10^8$	$1.6288 \times 10^7$	0.9437

cases has gradually decreased since the beginning of August. With only a few hundred cumulative cases, India has adopted prevention and control measures relatively early. Since a city

closure policy was implemented on March 25, 2020, the country went into emergency closure. Considering that India boasts a huge population and most of Indians lack enough literacy level and awareness of prevention and control, high-density aggregation of people have caused the virus to spread across the country, resulting in large  $c$ , small  $q$ , and small  $\rho$ , which promotes the outbreak of India's COVID-19. In the theoretical analysis, we simulate the development trend of the COVID-19 epidemic under different prevention and control measures by changing  $c$ ,  $q$ , and  $\rho$ , which are used to evaluate the impact of prevention and control measures on the development trend of the COVID-19 epidemic. This is shown in Figures 4–6.

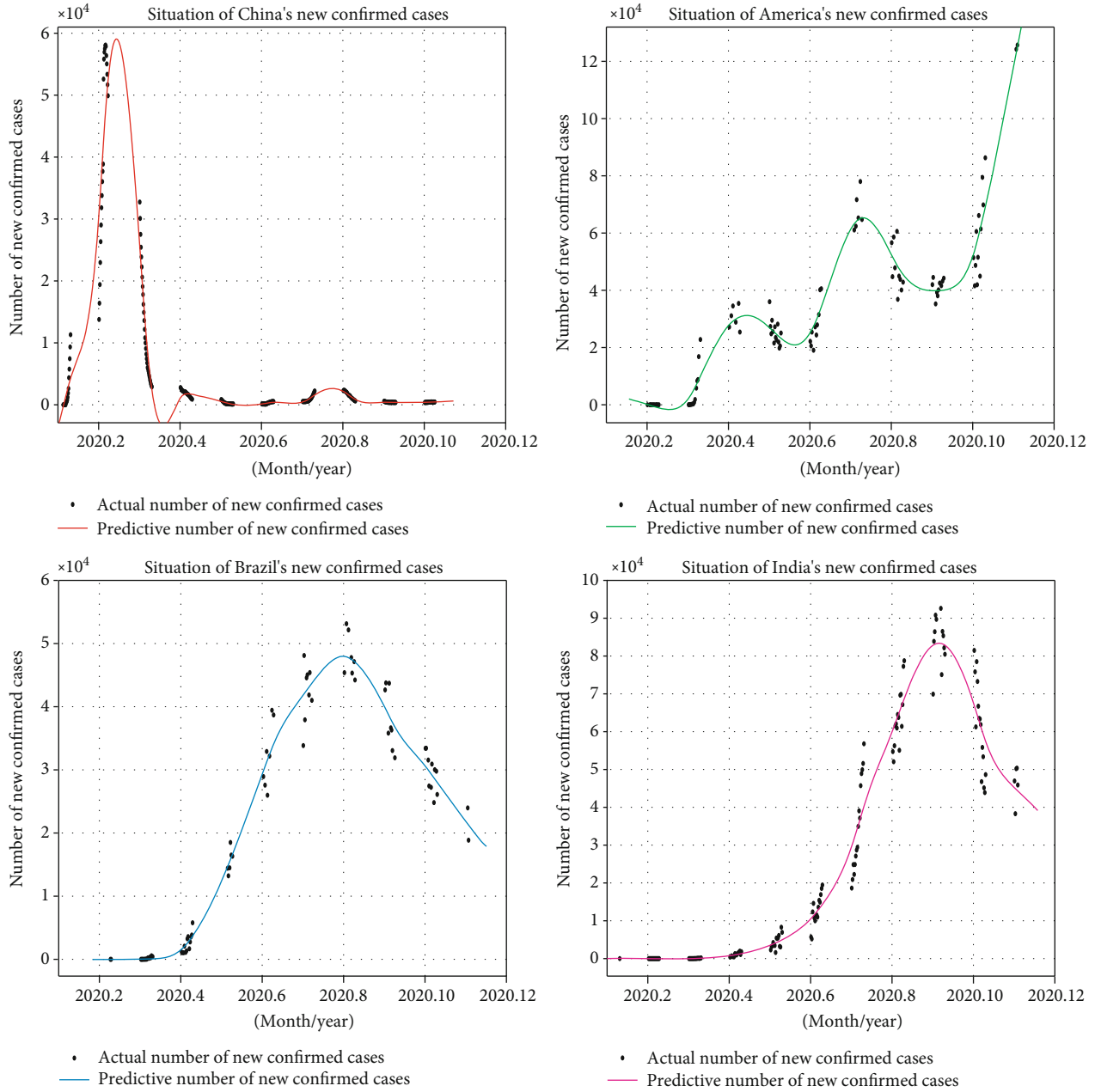


FIGURE 3: The fitting curve of the number of single-day confirmed cases.

As can be seen from Figure 4, when  $c$  is 1.5 times the actual situation in each country, the number of single-day confirmed cases in China will reach a peak of about  $7 \times 10^4$  in mid-February 2020. In comparison, China reached the earliest and the number of confirmed cases per day tends to be zero at the end of March and remains stable. In America, the number of single-day confirmed cases has continued to increase after reaching  $7 \times 10^4$ . The peak in Brazil has reached about  $6 \times 10^4$  and then gradually decreased. In India, the number of single-day confirmed cases reached  $7 \times 10^4$  in August. It will continue to increase to  $10 \times 10^4$  and then gradually decreased. When comparing the original curve ( $1.0q$ ) with the different values of the quarantine ratios for each country, it can be observed in Figure 5 that there is a sig-

nificant delay in the peak number of single-day confirmations at each stage in America, and the peak number of single-day confirmations in the other three countries is correspondingly earlier. The time is correspondingly advanced. As  $q$  increases, the peak number of confirmed cases in a single day decreases. When the quarantine ratio of each country is 0.5 times the actual situation, the number of single-day confirmed cases in China will reach a peak of about  $7 \times 10^4$  in February 2020 and stabilize in March. Around June 2020, the number of confirmed cases in America will reach  $7 \times 10^4$  and continue to increase to reach the second-stage peak of about  $8 \times 10^4$ . The number of COVID-19 confirmed cases in Brazil and India reached  $7 \times 10^4$  in a single day in August 2020, respectively, but both will continue to increase

TABLE 4: Peak number of confirmed countries in a single day when parameters change.

Value	Country	Parameter			
		$c$	$\rho$	$r$	$\rho$
0.25	China	32830	106160	110060	34530
	America	71140	233740	244150	66090
	Brazil	25570	105510	104060	33650
	India	53750	176190	181270	44490
0.5	China	45540	83200	83480	48340
	America	111020	179980	188650	102090
	Brazil	51580	88320	87040	50550
	India	83570	141810	135300	72440
1.0 (actual peak value)	China	58097	58097	58097	58097
	America	132540	132540	132540	132540
	Brazil	60091	60091	60091	60091
	India	97570	97570	97570	97570
1.5	China	78660	48820	45740	77580
	America	180110	103840	104610	185430
	Brazil	79290	53480	51260	83470
	India	140480	72470	81600	138350
2.0	China	107090	30910	34570	107590
	America	241630	62220	64350	235630
	Brazil	105480	25470	31070	108240
	India	177780	48770	45470	178010

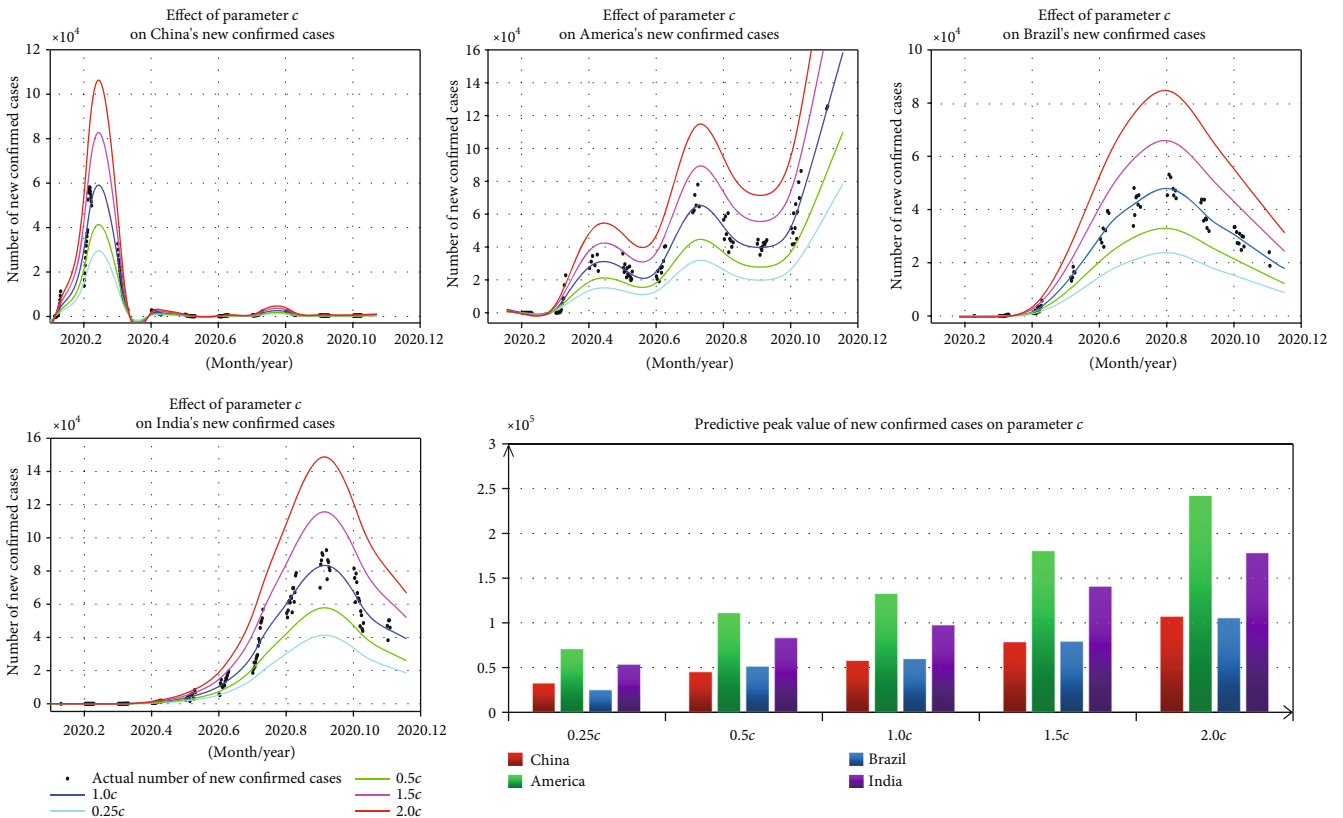


FIGURE 4: Impact of the contact rate on the COVID-19 outbreak.

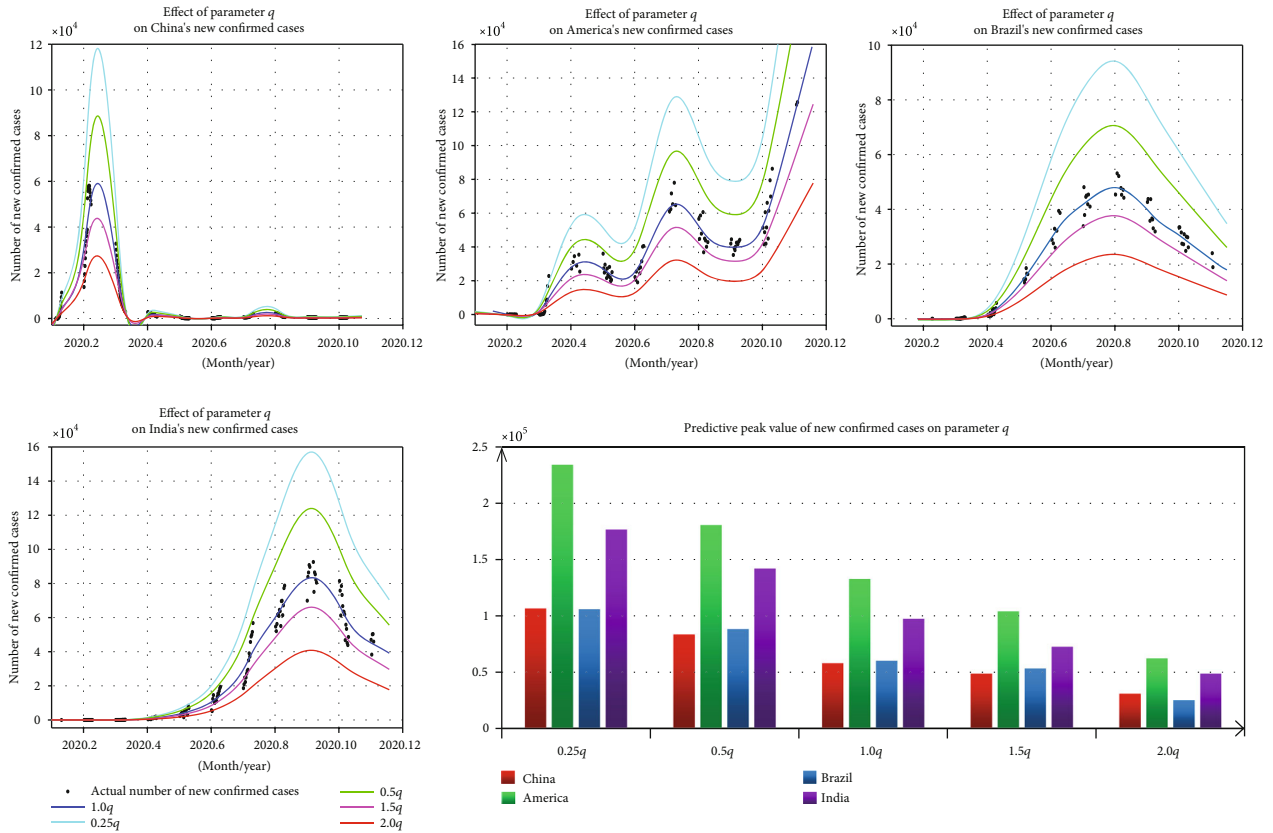


FIGURE 5: Impact of the quarantine ratio on the COVID-19 epidemic.

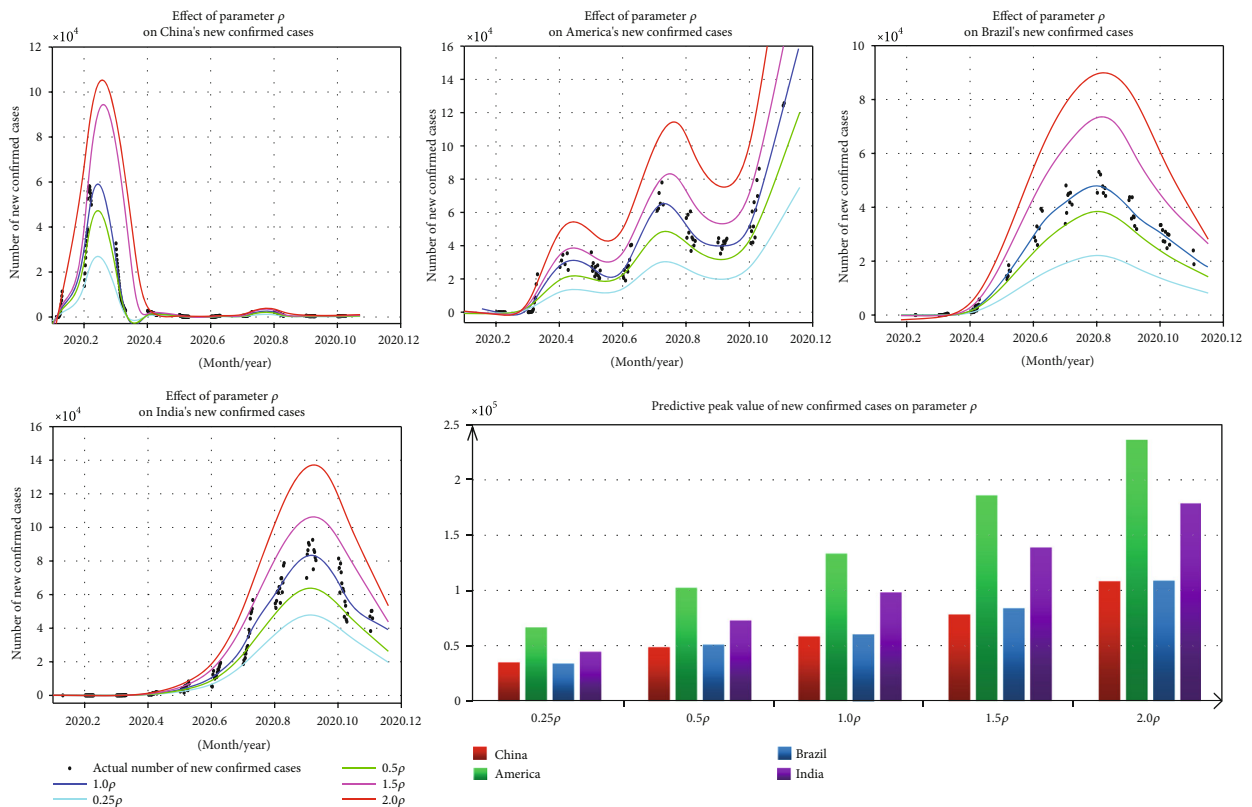


FIGURE 6: The impact of the effective exposure coefficient on the COVID-19 epidemic.

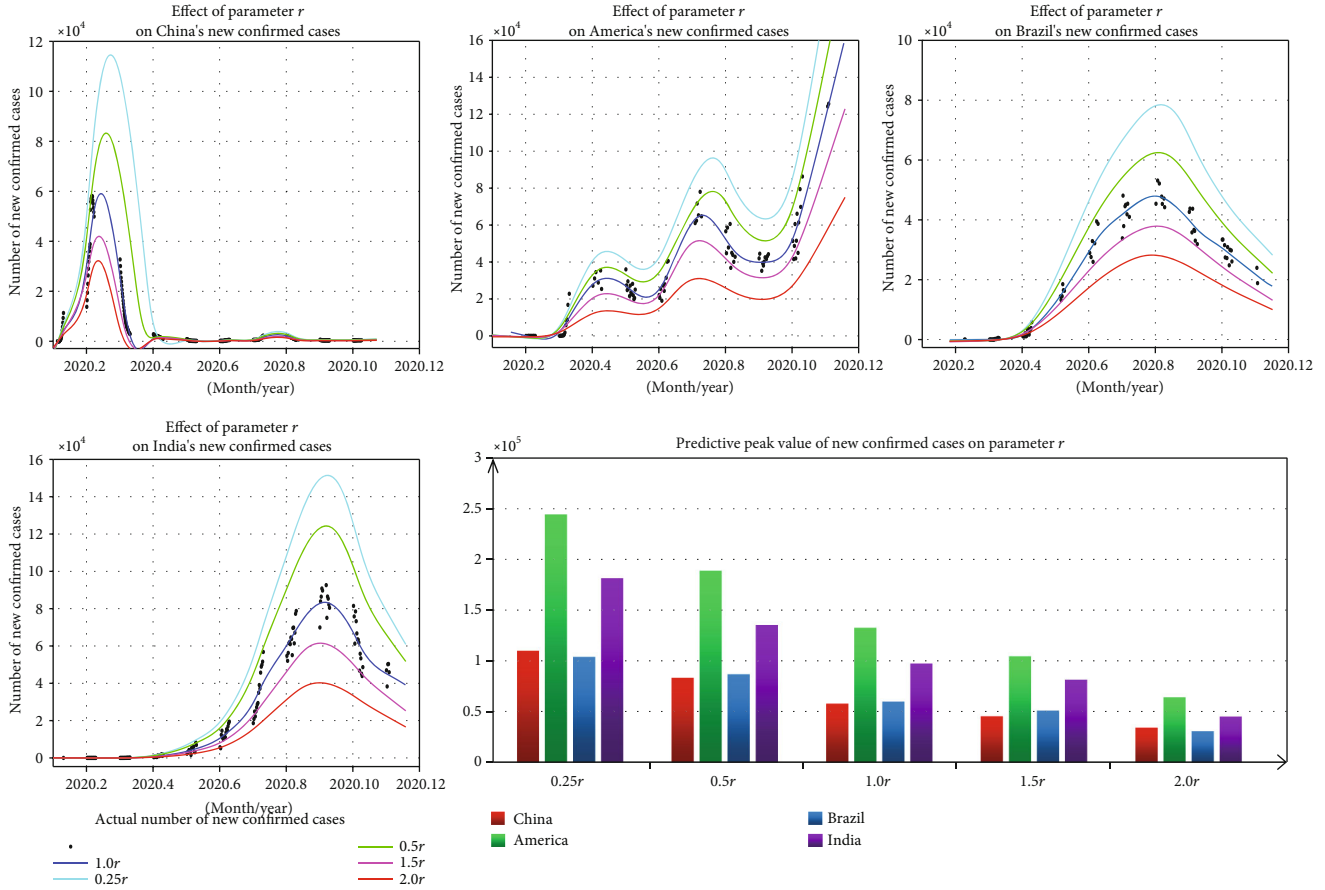


FIGURE 7: The impact of changes in the recovery rate on COVID-19.

to their respective peaks. According to Figure 6, at  $\rho$  of 0.25 times, it takes only one month for the number of confirmed cases in China to reach the highest peak of  $3 \times 10^4$  in a single day, while it takes only one month for Brazil and India to reach  $3 \times 10^4$  in mid-to-late July. They need to continue to increase for about a month before reaching their respective peaks. The second-phase peak of the epidemic in America was about  $3 \times 10^4$ . It is expected to continue to increase on a single day.

In China, the government has adopted greater and more rapid prevention and control measures. Chinese people have a strong awareness of prevention and control. So the development of COVID-19 reaches its peak fastest and is rapidly stable. In Brazil and India, the prevention and control measures and people’s awareness of prevention and control were relatively weak so that the COVID-19 epidemic reached its peak later. In the United States, the prevention and control measures and people’s awareness of prevention and control are weak and continue to increase after reaching the peak of the stage. It indicates that strict and timely prevention and control measures are taken. The higher the awareness of prevention and control, the more effectively the development of the COVID-19 epidemic can be suppressed.

5.2. Assessment of the Impact of Medical and Economic Levels on the COVID-19. As of May 31, 2020, China have allocated a

total of ¥162.4 billion in COVID-19 prevention and control at all levels of finance to ensure that financial support for COVID-19 prevention and control measures is in place. After the COVID-19 outbreak, the Chinese government has effectively increased the recovery rate through a large-scale free nucleic acid test, thereby ensuring receiving timely treatment of patients. America invested \$8.3 billion in the prevention and control of COVID-19 in the early stage of the COVID-19 outbreak. However, the number of single-day confirmed cases in America reached a peak after April 5, 2020, due to the limited testing capacity. To prevent the collapse of the medical system, the U.S. government adopted restrictions on testing, resulting in low  $r$ . Then, the number of single-day confirmed cases fluctuated up and down at the peak. Brazil’s uneven distribution of medical resources, insufficient reserves, and a fragile medical system have led to low  $r$ , which directly led to the acceleration of the increase in the number of single-day confirmed cases of the COVID-19 epidemic. India’s weak basic medical facilities and social medical security system, low virus detection capacity, small size of land area, high population density, and lack of medical equipment have led to the inability to treat infected patients in a timely manner, resulting in low  $r$ . This has led to a rapid increase in the number of single-day confirmed cases in June 2020. In the theoretical analysis, we simulate the development trend of the COVID-19 epidemic under different

medical and economic levels by changing  $r$  which is used to evaluate the impact of the medical level and economic level on the development trend of the COVID-19 epidemic, as shown in Figure 7.

It can be seen from Figure 7 that under the same scenario (2.0 times the actual  $r$ ), the COVID-19 epidemic in China will reach a peak of  $3 \times 10^4$  in early February 2020, with the number of single-day confirmed cases stabilizing around March. The number of single-day confirmed cases in Brazil will reach a peak of about  $2.5 \times 10^4$  in July 2020 and stabilize after December. The number of single-day confirmed cases of COVID-19 in India will reach  $3 \times 10^4$  in mid-April and then continue to increase to a peak of  $5 \times 10^4$ . As  $r$  in the three countries increases, the time to reach the peak of the number of single-day confirmed cases will be correspondingly advanced and the peak will be correspondingly smaller, with the most significant effect in India. Around July 2020, the number of single-day confirmed cases in America reached  $3 \times 10^4$ , which dropped slightly after reaching the peak of the second phase. However, the number of single-day confirmed cases increased significantly in October.

The Chinese government has invested heavily in the prevention and control of the COVID-19 epidemic and quickly adopted prevention and control measures. The COVID-19 epidemic has stabilized quickly. In Brazil and India, the number of single-day confirmed cases have peaked and stabilized due to fragile medical systems and other reasons. In the United States, although a large amount of money was invested in prevention and control, the uneven distribution of follow-up medical resources has led to a continuous increase in the number of single-day confirmed cases. It shows that increasing capital investment in epidemic prevention and control and sufficient medical resources are conducive to controlling the development trend of the COVID-19 epidemic.

## 6. Conclusion

In this paper, the transmission dynamic characteristics of COVID-19 are analyzed in four countries based on the SEIR-AQ model: China, America, India, and Brazil. As the economic level, medical level, prevention and control awareness, and prevention and control policies adopted by the four countries China, America, India, and Brazil are significantly different, we fully compare the differences in dynamic parameters such as  $c$ ,  $q$ ,  $r$ , and  $\rho$  among the above four countries and focus on the effects of prevention and control measures, awareness of prevention and control, economic level, and medical level on these parameters. A series of evidence shows that the control of the development of the COVID-19 epidemic requires the local government to quickly take prevention and control measures to minimize the number of exposed people in order to reduce the exposure rate and increase the isolation rate. On this basis, it should also increase financial investment in improving the economic level and medical level to increase the recovery rate, to realize the rapid conversion of confirmed patients to the recovered population. In addition, it is also crucial to raise awareness of prevention and control.

## Data Availability

The labeled dataset used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare no conflicts of interest.

## Acknowledgments

This work is supported by college students' Innovative Entrepreneurial Training Plan Program Grant No. 202010295015Z and the National Natural Science Foundation of China under Grant 61702225.

## References

- [1] L. Lanjuan and Z. Xueling, "Progress in the prevention and control of the COVID-19," *Zhejiang Medical Journal*, vol. 43, no. 1, pp. 1–8, 2021.
- [2] X. Jin, J. S. Lian, J. H. Hu et al., "Epidemiological, clinical and virological characteristics of 74 cases of coronavirus-infected disease 2019 (COVID-19) with gastrointestinal symptoms," *Gut*, vol. 69, no. 6, pp. 1002–1009, 2020.
- [3] L. L. R. Du RH, C. Q. Yang, W. Wang et al., "Reply to: Re: predictors of mortality for patients with COVID-19 pneumonia caused by SARS-CoV-2: a prospective cohort study," *The European respiratory journal*, vol. 55, 2020.
- [4] E. Beretta and Y. Takeuchi, "Global stability of an SIR epidemic model with time delays," *Journal of Mathematical Biology*, vol. 33, no. 3, pp. 250–260, 1995.
- [5] K. L. Cooke and P. van den Driessche, "Analysis of an SEIRS epidemic model with two delays," *Journal of Mathematical Biology*, vol. 35, no. 2, pp. 240–260, 1996.
- [6] Z. Yu, G. Zhang, Q. Liu, and Q. Lv, "The outbreak assessment and prediction of COVID-19 based on time-varying SIR model," *Journal of University of Electronic Science and Technology of China*, vol. 49, no. 3, pp. 357–361, 2020.
- [7] F. Wei, J. Wang, X. Xu et al., "Tendency prediction of COVID-19 worldwide," *Disease Surveillance*, vol. 35, no. 6, pp. 467–472, 2020.
- [8] H. Geng, A. Xu, X. Wang, Y. Zhang, X. Yin, and M. A. Mao, "Analysis of the role of current prevention and control measures in the epidemic of corona virus disease 2019 based on SEIR model," *Journal of Jinan University (Natural Science & Medicine Edition)*, vol. 41, no. 2, pp. 175–180, 2020.
- [9] W. Guozhu, C. Xiaohang, and Z. Qiang, "Forecast and analysis of epidemic situation based on improved SEIR model," *Journal of Henan Institute of Technology*, vol. 28, pp. 35–39, 2020.
- [10] S. Junjie, Y. Shixiong, G. Jingjing, Y. Ming, L. Zhong, and J. Nan, "Comparative analysis of the early transmission characteristics of COVID-19 epidemic between Shandong Province in China and South Korea based on the SEIR model," *Journal of Central China Normal University*, vol. 54, pp. 166–171, 2020.
- [11] L. Weiwei, D. Rong, C. Shudong, and S. Shuang, "Analysis of transmission characteristics of COVID-19 and prediction of the development trend of epidemic situation," *Journal of Xiamen University (Natural Science)*, vol. 59, no. 6, pp. 1025–1033, 2020.



- [12] J. Lin, "Assessment and prediction of COVID-19 based on SEIR model with undiscovered people," *Journal of University of Electronic Science and Technology of China*, vol. 49, no. 3, pp. 375–382, 2020.
- [13] C. Zhenyu, W. Junfen, Q. Kai et al., "Estimation of the COVID-19 epidemic in Italy by an SEAIQR model," *Modern Digestion & Intervention*, vol. 25, no. 3, pp. 273–279+283, 2020.
- [14] S. Ansumali, S. Kaushal, A. Kumar, M. K. Prakash, and M. Vidyasagar, "Modelling a pandemic with asymptomatic patients, impact of lockdown and herd immunity, with applications to SARS-CoV-2," *Annual Reviews in Control*, vol. 50, pp. 432–447, 2020.
- [15] B. Ivorra, M. R. Ferrández, M. Vela-Pérez, and A. M. Ramos, "Mathematical modeling of the spread of the coronavirus disease 2019 (COVID-19) taking into account the undetected infections. The case of China," *Communications in nonlinear science & numerical simulation*, vol. 88, p. 105303, 2020.
- [16] C. Pai, A. Bhaskar, and V. Rawoot, "Investigating the dynamics of COVID-19 pandemic in India under lockdown," *Chaos, solitons, and fractals*, vol. 138, 2020.
- [17] S. Cao, P. Feng, and P. Shi, "Study on the epidemic development of COVID-19 in Hubei Province by a modified SEIR model," *Journal of Zhe Jiang University (Medical Sciences)*, vol. 49, no. 2, pp. 178–184, 2020.
- [18] J. Jiao, Z. Liu, and S. Cai, "Dynamics of an SEIR model with infectivity in incubation period and homestead-isolation on the susceptible," *Applied Mathematics Letters*, vol. 107, p. 106442, 2020.

## Research Article

# Power Density Case Study for 5G mmWave Array Antennas

Dianyuan Qi <sup>1</sup>, Fangzhu Zou <sup>1</sup>, Jing Zhao <sup>1</sup>, Shaobin Sun <sup>2</sup>, Huanbin Wei <sup>2</sup>,  
Yiling Chen <sup>2</sup> and Zhan Xia <sup>2</sup>

<sup>1</sup>China Academy of Information and Communications Technology, Beijing 100191, China

<sup>2</sup>Huawei, Shenzhen 518129, China

Correspondence should be addressed to Fangzhu Zou; [zoufangzhu@caict.ac.cn](mailto:zoufangzhu@caict.ac.cn)

Received 25 January 2021; Revised 31 March 2021; Accepted 13 May 2021; Published 9 June 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Dianyuan Qi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

As 5G millimeter wave (mmWave) wireless device involves some new technologies, such as beamforming, the radiofrequency (RF) exposure compliance test for the 5G mmWave wireless device is significantly complicated. In order to shorten the compliance period for 5G mmWave terminals, the relevant regulatory authorities recommend a combination of numerical simulation and measurements to demonstrate compliance. To verify the feasibility of this method, the RF exposure test conducted in this paper was a reverse procedure according to IEEE (the Institute of Electrical and Electronics Engineers) standards. First, actual measurements under various conditions, including different beam configurations, different test distances, different input power levels, different duty cycle, and nonpeak directions, were performed, and the changing trend of PD over testing conditions was analyzed. Then one dual-polarized patch antenna array was selected for simulation analysis. The feasibility of the method proposed in IEEE standards was proved through the comparison of the results experiment and numerical analysis.

## 1. Introduction

The 3rd Generation Partnership Project (3gpp) protocol divides the 5G frequency spectrum into two FR (Frequency Range). FR1 is a low-frequency band up to 7125 MHz, and FR2 ranges from 24.25 GHz to 52.6 GHz. Most of the electromagnetic wavelengths in this FR2 are in the millimeter level. As mmWave technology can achieve a faster speed and lesser extent latency than 4G network, 5G wireless device has attracted massive attention in recent years. In the meantime, it also has raised the public concern regarding the biological effects on the human body resulted from the exposure to such high-frequency electromagnetic field (EMF). To control the effect in a safe range, standard organizations such as the International Commission for the Protection of Non-Ionizing Radiation (ICNIRP) and IEEE have established the permissible exposure limits for electromagnetic radiation [1].

Specific absorption rate (SAR) and power density (PD) are usually adopted to determine exposure compliance. SAR defines the amount of energy absorbed by human tissue per mass unit, and it is currently preferred in exposure com-

pliance determination. However, when higher frequencies are involved, the skin depth decreases, and energy absorption is usually limited to the skin surface. Studies have shown that when the frequencies are at 10 GHz-100 GHz, about 90% of the energy will be absorbed by the skin's epidermis or dermis [2]. In such circumstance, it is difficult to build a meaningful volume for SAR. The mmWave devices operating at such high-frequency range should be evaluated with power density, as it is expressed per unit area for near-field RF exposure evaluation, instead of per unit mass in SAR. Whereas power density measurement is a time-consuming procedure; therefore, combining simulation and measurement has been proposed as an efficient method during a compliance test.

## 2. Device under Test

The device used in this testing is a prototype board that operates at 28 GHz, and the bandwidth is 200 MHz. There are five 5G antenna arrays at different sides of the device as shown in Figure 1. The details of the device antenna are as follows:

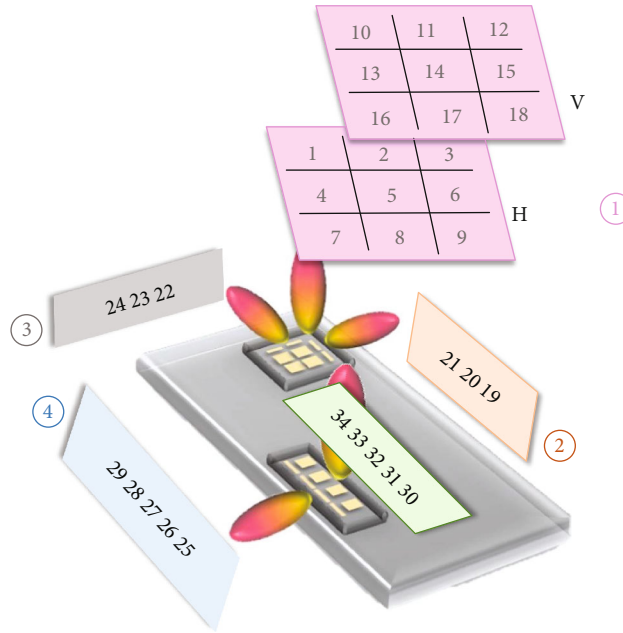


FIGURE 1: Antenna location and beam configuration (rear view).

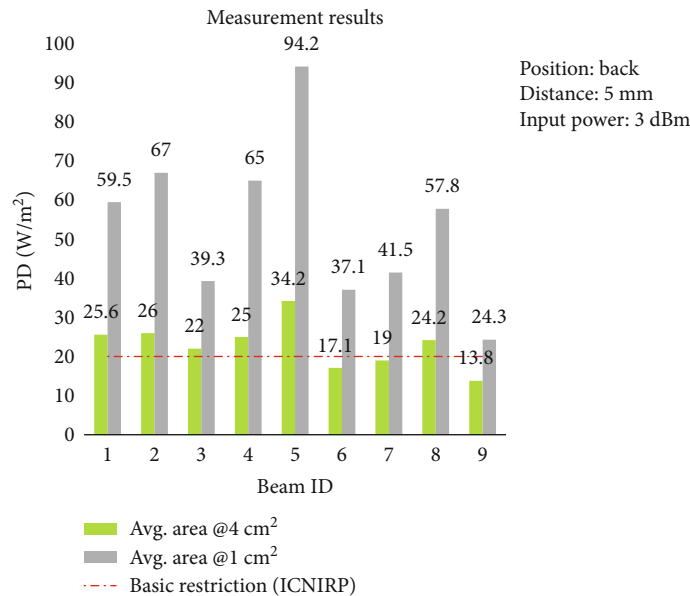


FIGURE 2: Measurement results of beam ID 1-9.

- (1) One  $2 \times 2$  patch antenna array, dual-polarized, is located at the left corner of DUT. Beam 1-18 are generated by this dual-polarized patch antenna, of which beam 1-9 are horizontal polarization and beam 10-18 are vertically polarized
- (2) Two  $1 \times 2$  dipole antenna arrays, single-polarized, are located at the left corner of the device. Beam 19-20 are generated by the antenna at the left edge. Beam 22-24 are generated by the antenna at the top edge
- (3) One  $1 \times 4$  dipole antenna array, single-polarized, is located at the right edge of the device. Beam 25-29 are generated by this antenna
- (4) One  $1 \times 4$  patch antenna array, dual-polarized, is located at the right edge of DUT. Beam 30-34 are generated by this patch antenna

The antenna arrays cannot work simultaneously. While one antenna is active, the rest of the arrays will be disabled. As for beamforming, the beam configuration of an antenna array can be changed by software.

### 3. Power Density Assessment

The power density testing employs the mmWave module of cDASY6 from SPEAG, Switzerland. The mmWave

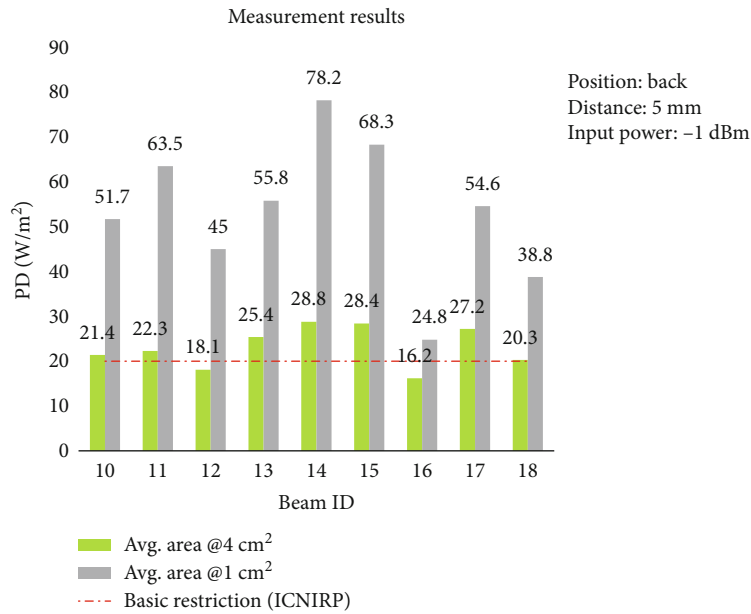


FIGURE 3: Measurement results of beam ID 10-18.

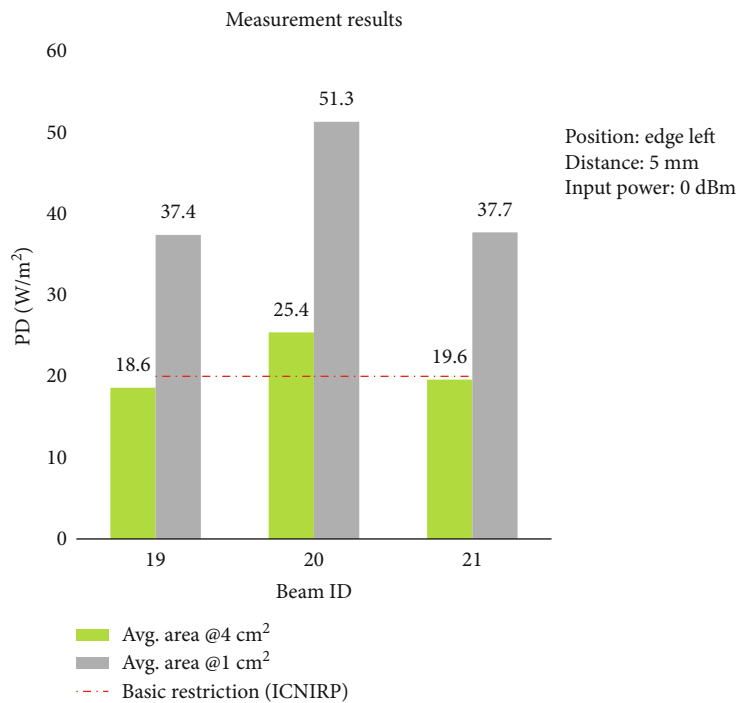


FIGURE 4: Measurement results of beam ID 19-21.

module supports a measurement at distances as small as 2 mm and to automatically determine the optimal measurement grid based on the frequency modulation, evaluation distance, and extent of the transmitter array.

The phantom of this mmWave module approximates free-space conditions and allows evaluating not only the antenna side of the device under test (DUT) but also any opposite-radiating side of the devices operating above 10 GHz.

The probe used in the measurement is an E-field mmWave probe. It can perform at the frequency range from 750 MHz to 110 GHz. According to the current draft of IEC/IEEE 63195, for general near-field measurements, the magnitudes and phases of both E-field and H-field are necessary to correctly determine the power density. As an E-field probe is adopted during measurement, the field reconstruction is used to derive the H-field from the direct measurement of E-field. All components of the E-field, the H-field,

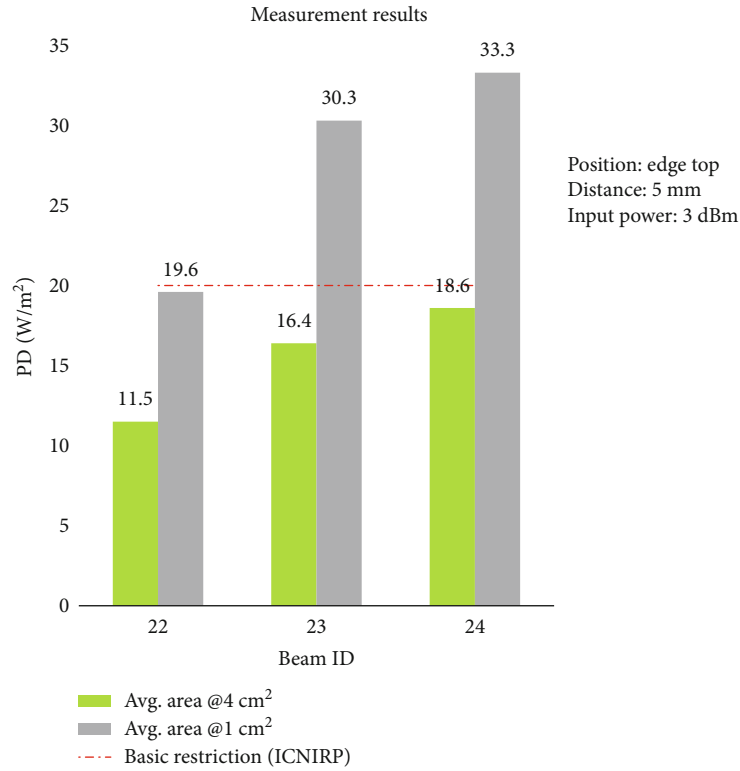


FIGURE 5: Measurement results of beam ID 22-24.

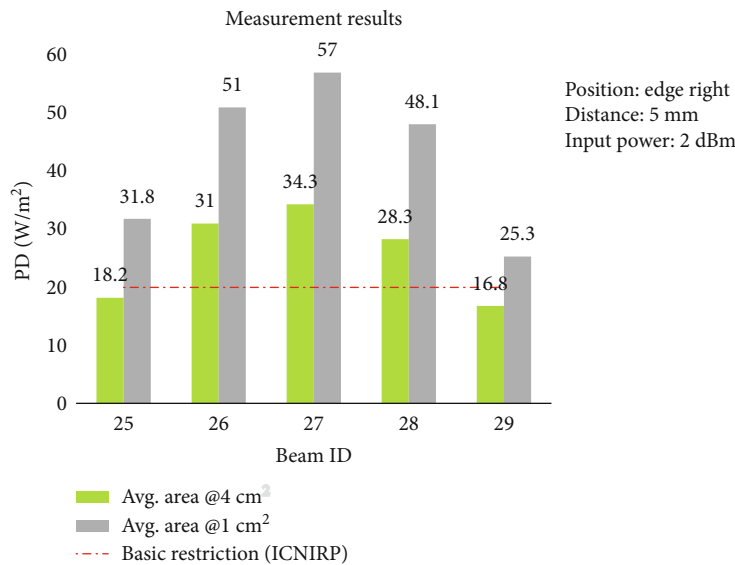


FIGURE 6: Measurement results of beam ID 25-29.

and the power density on the measurement plane are reconstructed through an approach based on the Gerchberg-Saxton algorithm with the mmWave module of cDASY6 [2, 3].

#### 4. Test Settings

During the power density measurement, the mmWave signal in the DUT is configured by an engineering testing software,

and the test mode is nonsignaling. This testing is implemented in 5 different situations:

- (1) The device is operated to transmit continuously at 100% transmission duty with different antenna beams. The test distance is 5 mm, and the input power is specified for each antenna array
- (2) For certain beams, 3 different test distances are set from 10 mm to 20 mm with 5 mm increments. The

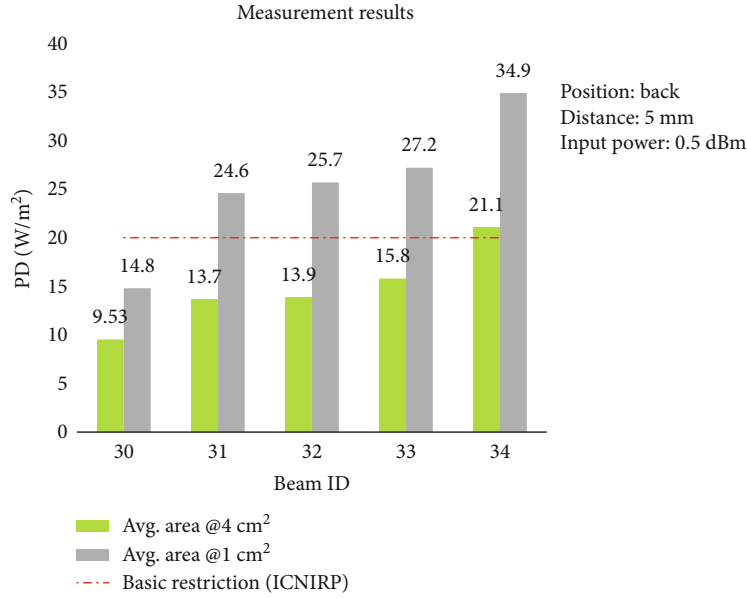


FIGURE 7: Measurement results of beam ID 30-34.

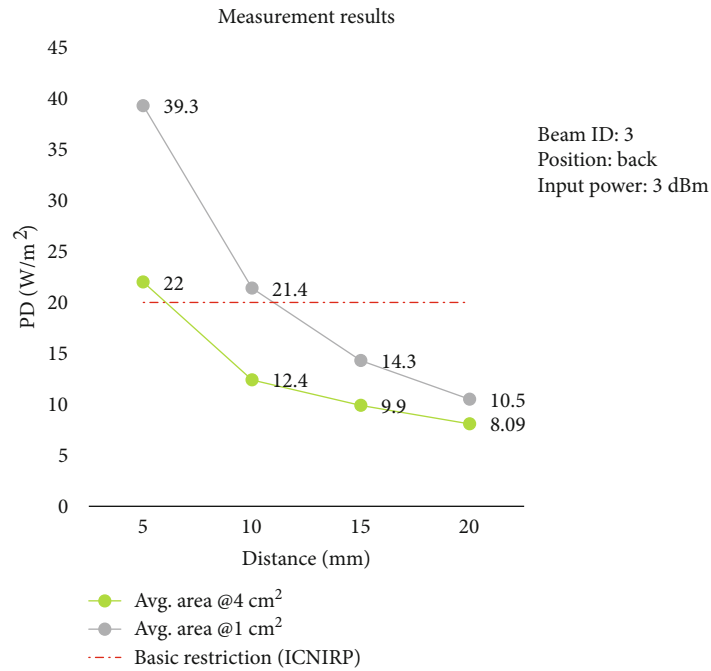


FIGURE 8: Measurement results at different test distances for beam 3.

device is operated to transmit continuously at 100% transmission duty, and the input power is specified for each antenna array

- (3) Each measurement of certain beams is performed with different input power levels and at 100% transmission duty. The test distance is 5 mm
- (4) The measurements of a certain beam are implemented in duty cycle of 20% and 60%, respectively, with 5 mm test distance

- (5) The measurements are implemented on the nonpeak directions of certain beams with specified input powers and 5 mm test distance

### 5. Measurement Results

The PD shown as follows is computed under the averaging area of 1 cm<sup>2</sup> and 4 cm<sup>2</sup>. As the antenna arrays operate at 28 GHz, according to ICNIRP RF Guidelines, the data of concern is averaged over 4 cm<sup>2</sup> in this paper.

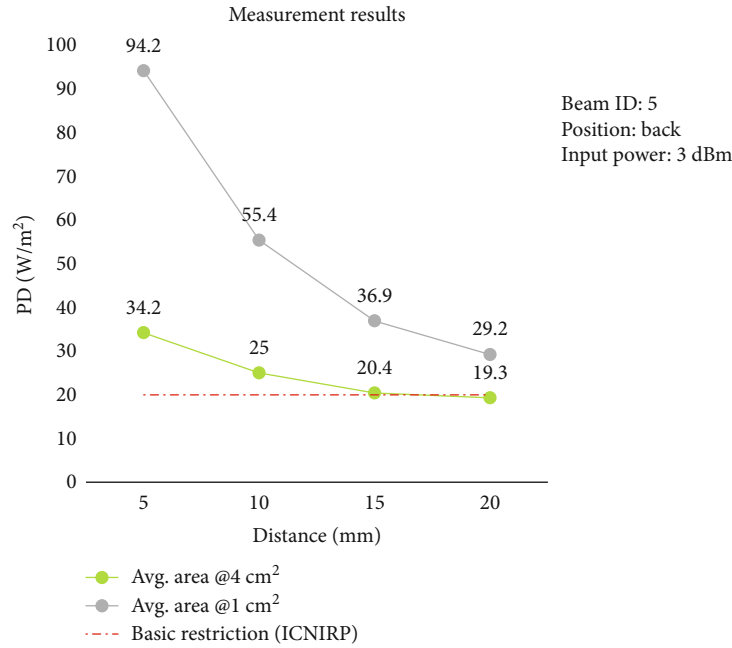


FIGURE 9: Measurement results at different test distances for beam 5.

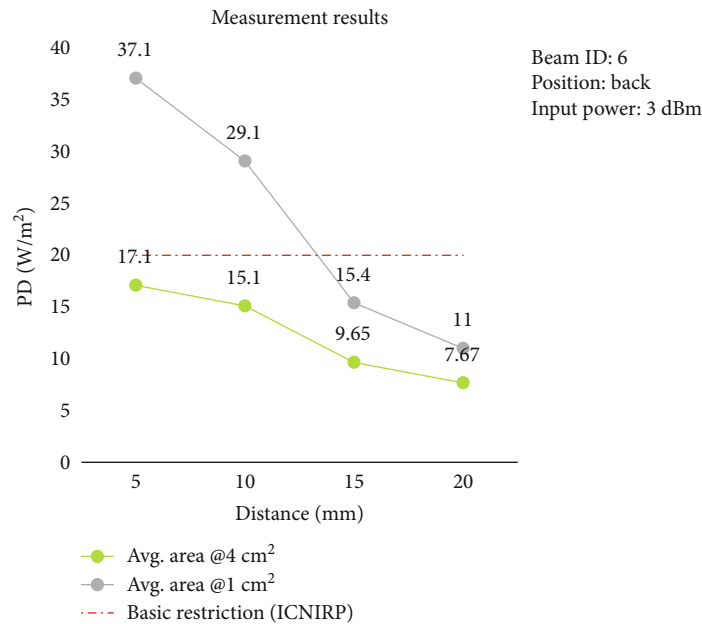


FIGURE 10: Measurement results at different test distances for beam 6.

5.1. *Power Density Measurement on Different Beam Configurations.* As shown in Figures 2 and 3, the highest measured power density beam ID is 5 at  $34.2 \text{ W/m}^2$ , and beam 14 is the maximum measured beam (power density of  $28.8 \text{ W/m}^2$ ), which are the central vectors for their respective polarizations. Moreover, the H polarization beams of this  $2 \times 2$  patch antenna have higher power density than its V polarization beams. Most of the values obtained from this antenna array are above the basic

restriction of  $20 \text{ W/m}^2$  as averaged over  $4 \text{ cm}^2$  according to ICNIRP.

Most of the beams of the single-polarized antenna array at the left edge are below the exposure limits, and the maximum power density is  $25.4 \text{ W/m}^2$  found from beam 20 (as shown in Figure 4).

Results in Figure 5 demonstrate that the PD of the antenna at the top edge does not exceed the current ICNIRP basic limit of  $20 \text{ W/m}^2$  at  $4 \text{ cm}^2$ .

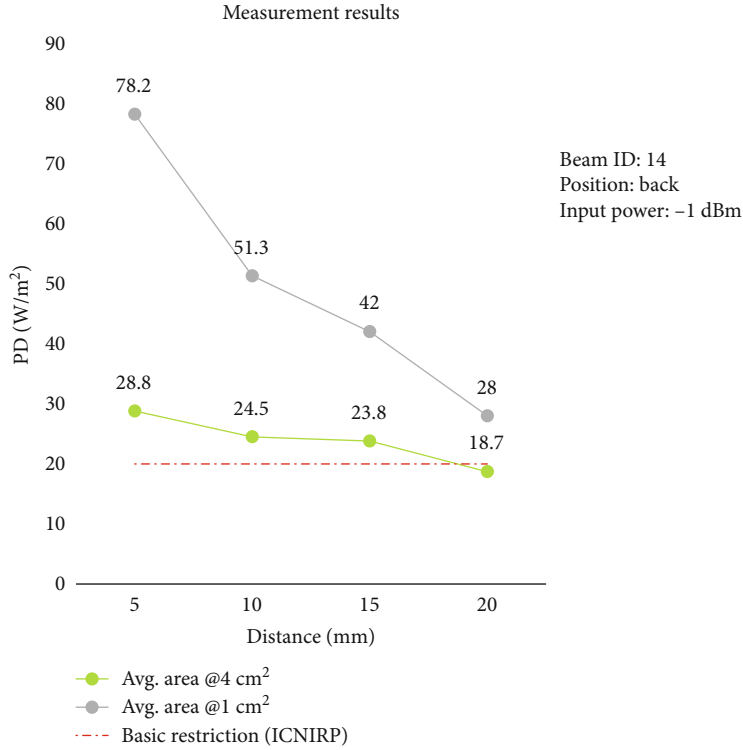


FIGURE 11: Measurement results at different test distances for beam 14.

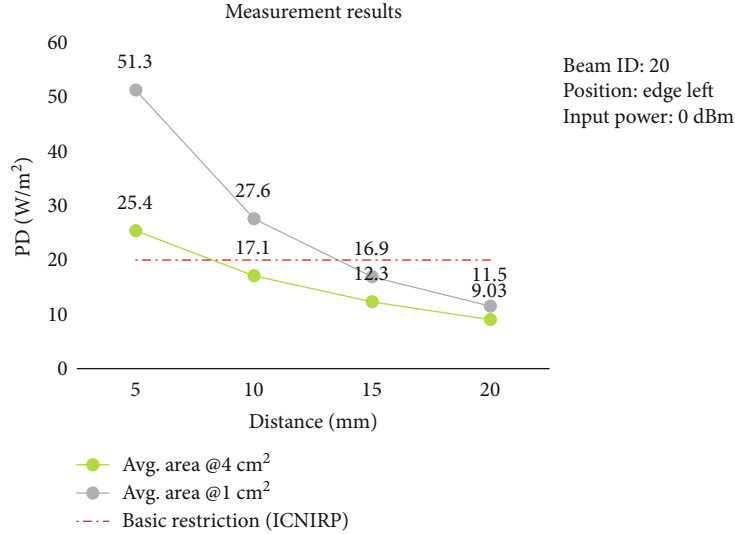


FIGURE 12: Measurement results at different test distances for beam 20.

The results shown in Figure 6 are measured from the single-polarized antenna array at the right edge. Out of the total five beams, three beams are larger than  $20 \text{ W/m}^2$ . The largest value is measured from beam 27 at  $34.3 \text{ W/m}^2$ .

Figure 7 is the measured results of the  $1 \times 4$  patch antenna array beams. They indicate that beam 30-33 comply with the RF radiation exposure limits of the ICNIRP.

*5.2. Power Density Measurement at Different Test Distances.* Data from Figures 8–15 show that the power density decreases as the testing distance increases from 5 mm to 20 mm. For test distance at 10 mm, the PD of beam 3, beam 20, and beam 27 complies with the limits set by ICNIRP. Figures 9 and 11 demonstrate that the PD of the two normal beams of the  $2 \times 2$  dual-polarized patch antenna array is still over the basic restriction even at the 15 mm test distance.



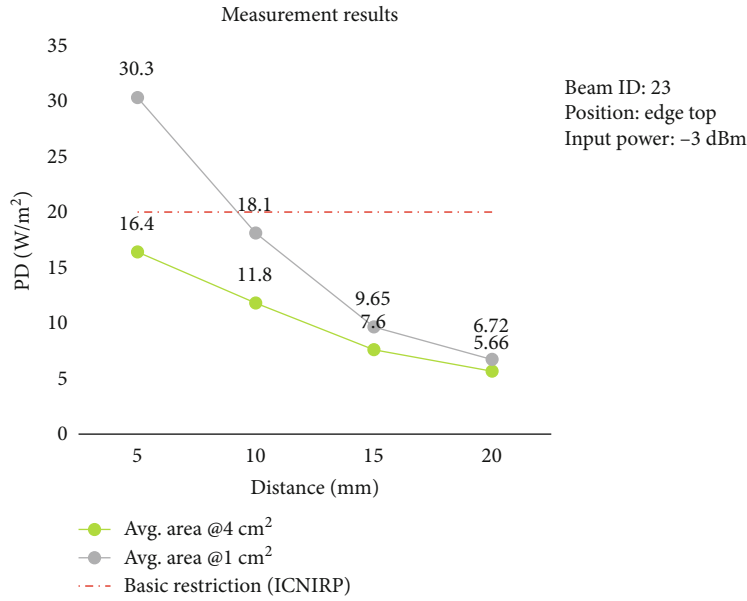


FIGURE 13: Measurement results at different test distances for beam 23.

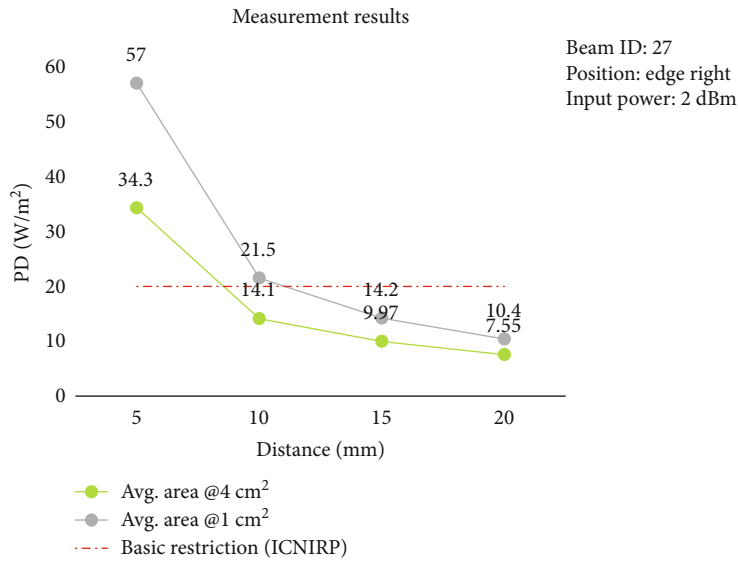


FIGURE 14: Measurement results at different test distances for beam 27.

5.3. Power Density Measurements on Different Input Power.

Figures 16 and 17 show that the measurement values go down when the input power levels are reduced by -3 dB and -5 dB, respectively, for both beam 14 and 23. The power density of beam 14 is finally below 20 W/m<sup>2</sup> at 4 cm<sup>2</sup> as the input power is decreased from -1 dB to -4 dB.

5.4. Power Density Measurements under Different Duty Cycle.

As shown in Figure 18, there is a significant trend in the measurement result change under different duty cycle, i.e., when the duty cycle is 0.6, the measured power density is approxi-

mately 60% of that at its maximum duty cycle, and the same is true when the duty cycle is 20%.

5.5. Power Density Measurements at the Nonpeak Direction.

Beam 20 is generated by the single-polarized dipole antenna located at the left corner. Figure 19 shows that the power density at the back (nonpeak direction) is close to the highest simulated power density at the left edge, while the power density of beam 23 generated by the left corner antenna at the back (nonpeak direction) is approximately equal to the highest power density at the top edge (as shown in Figure 20).

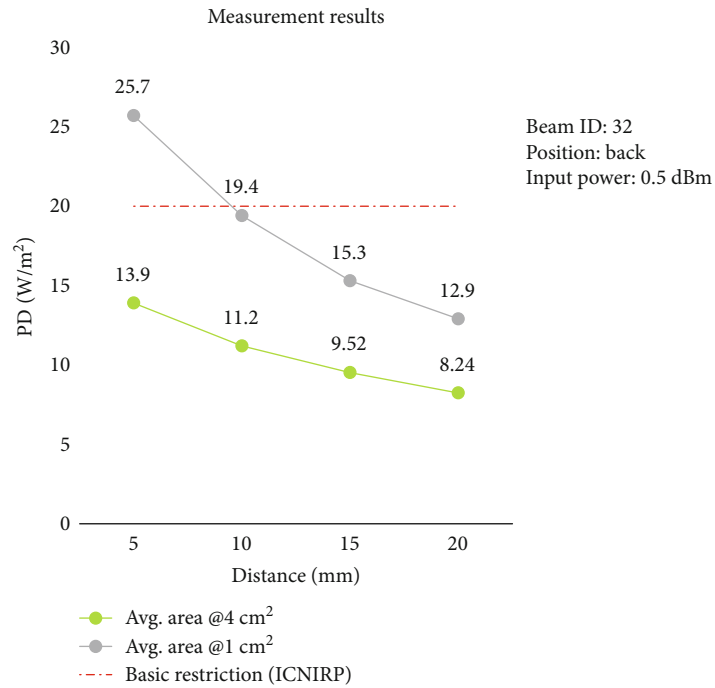


FIGURE 15: Measurement results at different test distances for beam 32.

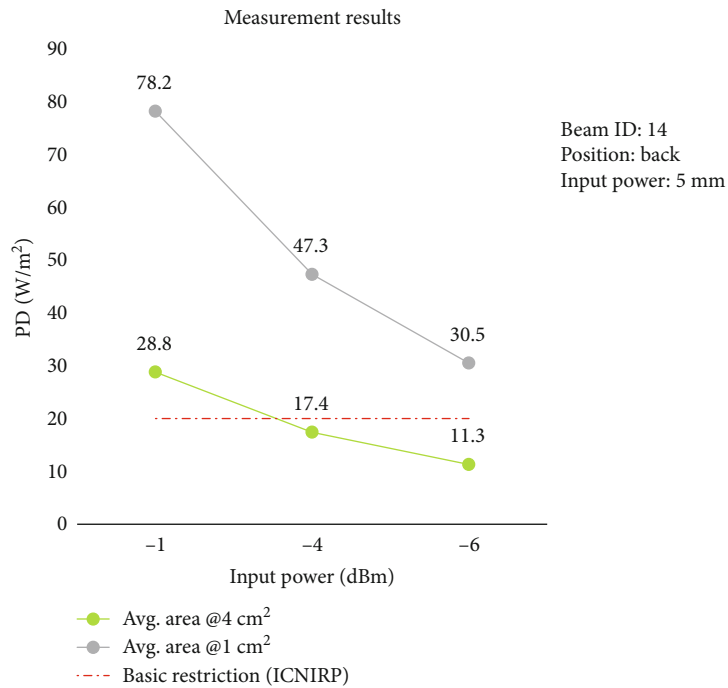


FIGURE 16: Measurement results on different input power for beam 14.

Beam 2, beam 6, and beam 8 are generated by  $2 \times 2$  patch antenna array that is located at the back. As can be seen from Figures 21–23, the measured power density of each beam at the nonpeak directions is much less than that on the major radiation direction.

## 6. Simulation and Measurement Results Comparison

6.1. *Simulation Environment.* For providing case study for power density analysis, we perform simulation and

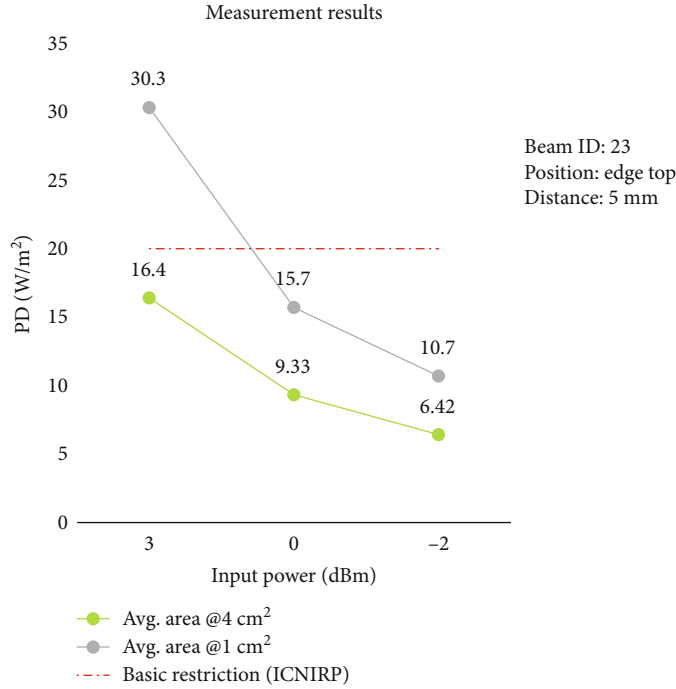


FIGURE 17: Measurement results on different input power for beam 23.

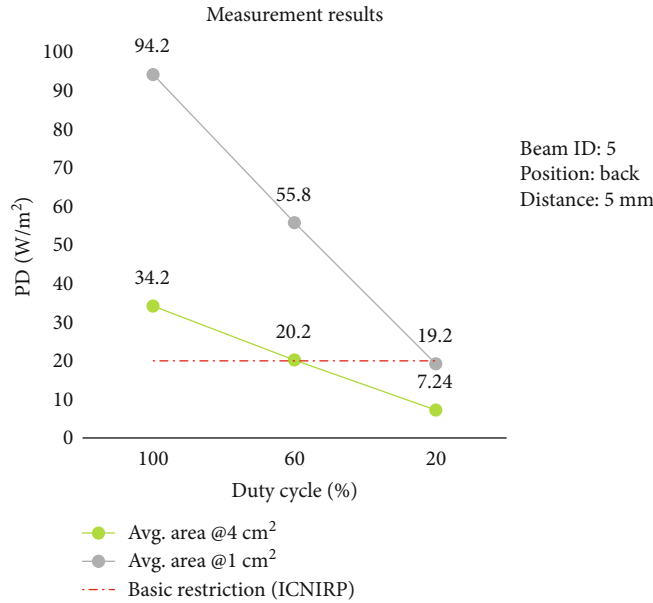


FIGURE 18: Measurement results under different duty cycle for beam 5.

measurement comparison for considered array antenna. The commercial simulation tool CST based on FDTD (Finite-Difference Time-Domain) algorithm is used for power density simulation.

To explore the consistency of simulation and measurement, the simulated results of  $2 \times 2$  patch antenna array will be discussed.

**6.2. Power Density Simulation Results.** In order to save computational resources, the DUT model is truncated and only

the  $2 \times 2$  patch antenna array located at the left corner is modeled and simulated in this paper.

The detailed simulation results of the selected  $2 \times 2$  patch antenna array beams are shown in Table 1. The DUT is modeled in free-space and the same configurations as the (1) testing condition mentioned in chapter 4, where the averaging area is  $4 \text{ cm}^2$  and distance is 5 mm from the back side.

The simulation results have been normalized to the same PA output power level for comparison with the measurement

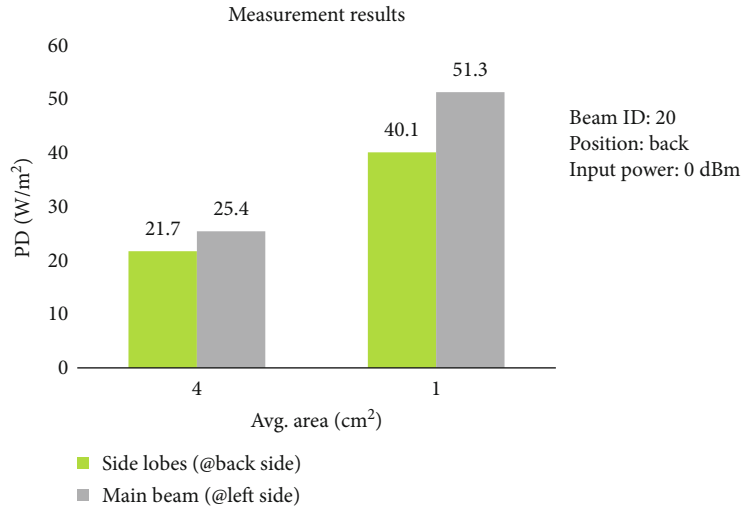


FIGURE 19: Measurement results in a nonpeak direction for beam 20.

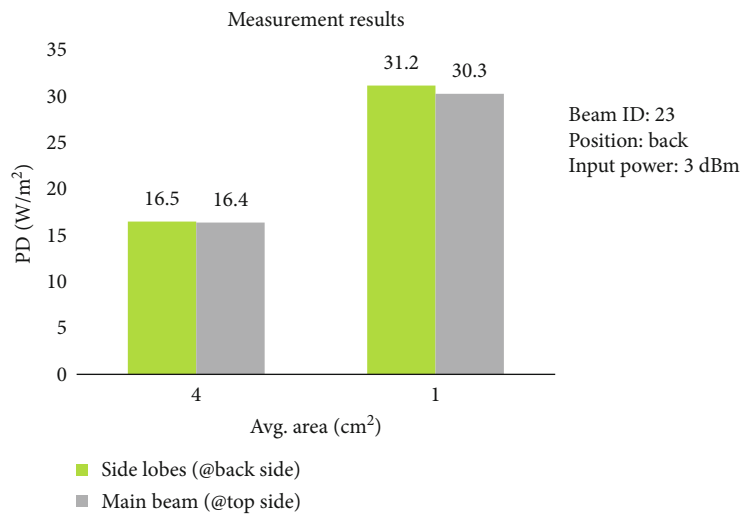


FIGURE 20: Measurement results in a nonpeak direction for beam 23.

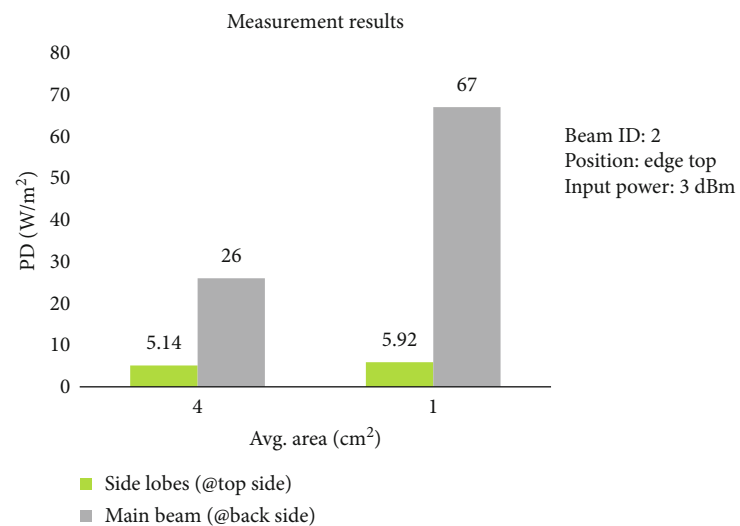


FIGURE 21: Measurement results in a nonpeak direction for beam 2.

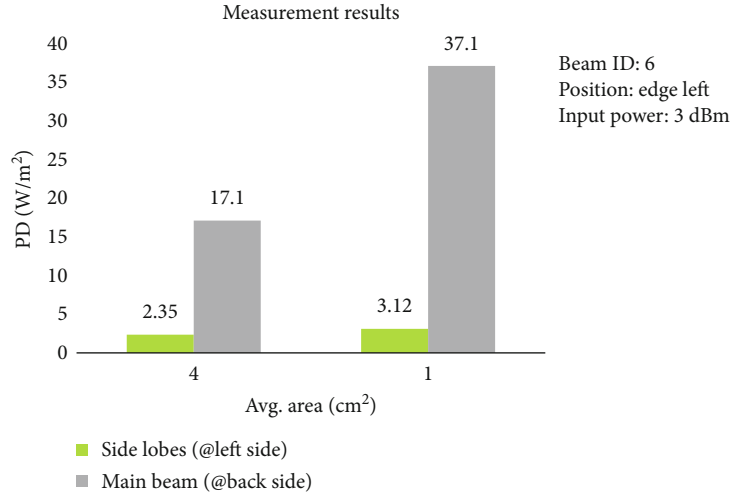


FIGURE 22: Measurement results in a nonpeak direction for beam 6.

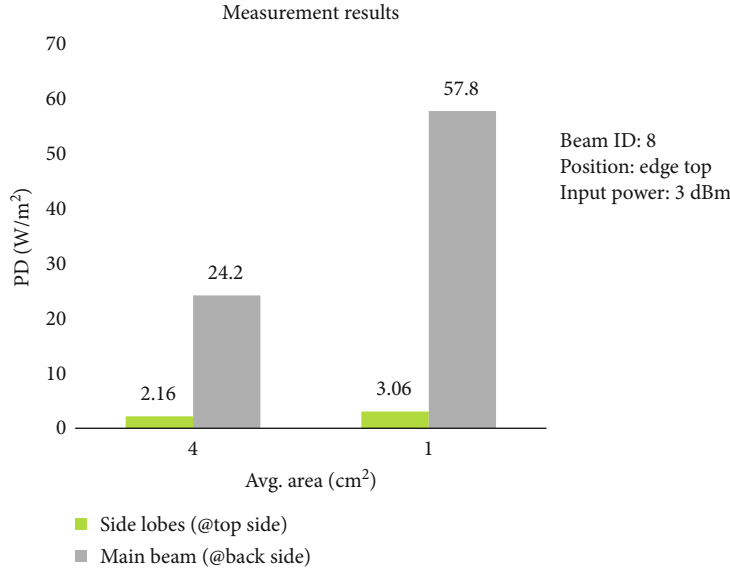


FIGURE 23: Measurement results in a nonpeak direction for beam 8.

results. The insertion loss difference of each panel has also been considered for the simulation and measurement results normalization.

6.3. Analysis

- (1) The deviations between PD simulation and measurement results of the selected  $2 \times 2$  patch antenna array beams were verified to be within  $\pm 2.4$  dB
- (2) From simulation and measurement results, we could check the correlation of power density variation. The PD trends between simulation and measurement may not ideal one-to-one correspondence. But the top four highest beams from PD

simulation at most can still cover the worst case of PD measurement

Figure 24 shows the cross-sectional power density plot of the normal and offset beams. The PD measurement results show that the normal beam (e.g., beam 5) is the worst case in the panel. While in the simulation, the PD simulation result of the offset beam (e.g., beam 4) is slightly higher than the normal beam.

In the near field simulation region, there may be a case where the offset beam PD is slightly higher than the normal beam due to the influence of the shell and the close distance. The different array elements may occur in the beam aggregation region. It is probably because the truncated DUT model is used, and the surrounding shells are not accurately modeled.

TABLE 1: PD simulation and measurement results comparison.

Test condition	Beam ID	Panel insertion loss (dB)	Normalized results		Deviation (dB)
			Simulation (W/m <sup>2</sup> )	Measurement (W/m <sup>2</sup> )	
2 × 2 patch antenna array, back side, 5 mm	1	0.6	31.72	25.6	-0.93
	2	0.6	25.71	26	0.05
	3	0.6	23.58	22	-0.3
	4	0.6	34.37	25	-1.38
	5	0.6	30.22	34.2	0.54
	6	0.6	25.88	17.1	-1.8
	7	0.6	31.72	19	-2.23
	8	0.6	25.69	24.2	-0.26
	9	0.6	23.68	13.8	-2.35
	10	1.45	23.22	21.4	-0.35
	11	1.45	24.99	22.3	-0.49
	12	1.45	23.17	18.1	-1.07
	13	1.45	29.89	25.4	-0.71
	14	1.45	31.06	28.8	-0.33
	15	1.45	31.25	28.4	-0.42
	16	1.45	23.42	16.2	-1.6
	17	1.45	25.76	27.2	0.24
	18	1.45	23.22	20.3	-0.58

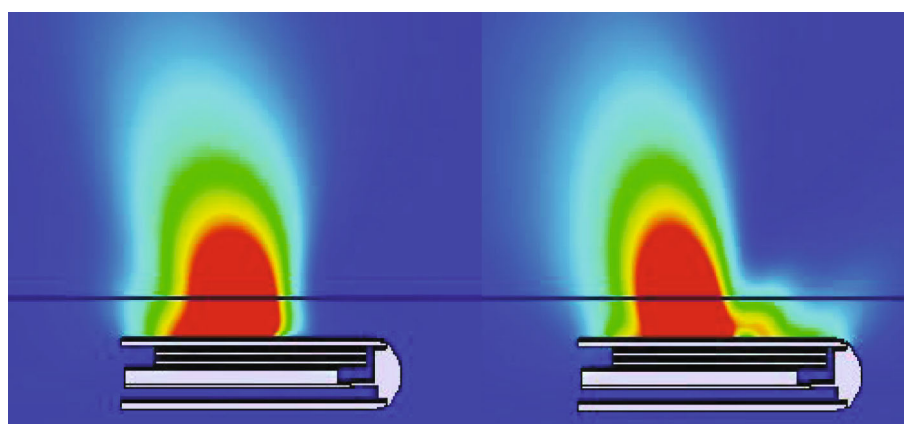


FIGURE 24: Cross-sectional power density plot of the normal and offset beams.

## 7. Conclusion

For mmWave array antennas, when practical, the combination of numerical simulation and measurements to demonstrate the power density compliance is a practicable assessment approach and a considerable reduction in the measurement period. As

- (1) from power density measurement results of multiple factors (different beam configurations, test positions, input power levels, duty cycles, main lobe and side lobe beam directions, etc.), we could check the correlation of power density variation. It would be a good reference when developing IEC standard

- (2) from power density simulation and measurement results comparison, it is worth to consider the power density results from the simulation in term of showing compliance

## Data Availability

The measurement and simulation data used to support the findings of this study are included within the article.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

- [1] K. R. Foster, S. Kodera, and A. Hirata, “5G communication systems and radiofrequency exposure limits,” *IEEE Future Networks Tech Focus*, vol. 3, no. 2, 2019.
- [2] T. Wu, T. S. Rappaport, and C. M. Collins, “The human body and millimeter-wave wireless communication systems: interactions and implications,” in *2015 IEEE International Conference on Communications (ICC)*, pp. 2423–2429, London UK, 2015.
- [3] American National Standards Institute, *Safety Levels with Respect to Human Exposure to Radio Frequency Electromagnetic Fields, 300kHz to 100GHz*, vol. 10016, The Institute of Electrical and Electronic Engineering New York, 1982.

## Research Article

# Association Analysis of Private Information in Distributed Social Networks Based on Big Data

**Dongning Jia** <sup>1,2</sup> **Bo Yin** <sup>1,2</sup> and **Xianqing Huang** <sup>2</sup>

<sup>1</sup>Ocean University of China, Qingdao, Shandong 266100, China

<sup>2</sup>Pilot National Laboratory for Marine Science and Technology (Qingdao), Qingdao, Shandong 266237, China

Correspondence should be addressed to Bo Yin; [ybfirst@ouc.edu.cn](mailto:ybfirst@ouc.edu.cn)

Received 29 April 2021; Revised 19 May 2021; Accepted 21 May 2021; Published 7 June 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Dongning Jia et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

As people's awareness of the issue of privacy leakage continues to increase, and the demand for privacy protection continues to increase, there is an urgent need for some effective methods or means to achieve the purpose of protecting privacy. So far, there have been many achievements in the research of location-based privacy services, and it can effectively protect the location privacy of users. However, there are few research results that require privacy protection, and the privacy protection system needs to be improved. Aiming at the shortcomings of traditional differential privacy protection, this paper designs a differential privacy protection mechanism based on interactive social networks. Under this mechanism, we have proved that it meets the protection conditions of differential privacy and prevents the leakage of private information with the greatest possibility. In this paper, we establish a network evolution game model, in which users only play games with connected users. Then, based on the game model, a dynamic equation is derived to express the trend of the proportion of users adopting privacy protection settings in the network over time, and the impact of the benefit-cost ratio on the evolutionarily stable state is analyzed. A real data set is used to verify the feasibility of the model. Experimental results show that the model can effectively describe the dynamic evolution of social network users' privacy protection behaviors. This model can help social platforms design effective security services and incentive mechanisms, encourage users to adopt privacy protection settings, and promote the deployment of privacy protection mechanisms in the network.

## 1. Introduction

Online social networks are changing people's daily behaviors, bringing great convenience to people's lives. With the frequent occurrence of privacy leaks, users have paid more and more attention to protecting the security of personal social data; the advent of the era of big data makes users' privacy and security face more threats. Therefore, the privacy and security issues of social networks have become a hot spot for users, service providers, and researchers [1, 2]. Relevant research work has proposed many methods and technologies to solve the problem of privacy leakage [3, 4]. Although it can alleviate this problem to a certain extent, it cannot completely eliminate the problem of privacy leakage. The main reason is that these studies have neglected the social network service provider's response to users. The root of this threat is the centralized social network service structure. In the centralized

service model, online social network service providers are the core of the entire system architecture, and users are deprived of the right to control personal data. All user data is exposed to social network service providers, which provides necessary conditions for service providers to collect user data [5].

Privacy protection is different from traditional access control technology and encryption technology in that it does not arbitrarily cut off the access channel of secret data, nor does it simply decode the data. With the rapid changes in network technology, the scope of sensitive data content is also changing, from the table structure in the earliest relational model to the later streaming data and social network data containing sensitive data. Social network has a huge user group and data volume. Therefore, social network has become a hot research object in many disciplines such as computer science, sociology, and psychology, and social



network analysis has also become an important branch of Web data mining [6]. The development of social networks has made relevant data sets easier to obtain, and the development of social network analysis has also increased the possibility of data privacy information leakage. Therefore, the availability of data and the privacy of information are the trade-offs for data release. Therefore, it is very necessary to study the privacy protection of social networks.

Based on the observation and analysis of the traditional differential privacy protection mechanism, this paper verifies the effectiveness of differential privacy and analyzes the necessity and infeasibility of differential privacy on social networks. The subject designed a differential privacy mechanism based on interactive social networks, clarified in detail the advantages of improved interactive differential privacy compared to traditional differential privacy and its feasibility on social networks, and conducted a strong verification of its effectiveness. It focuses on the algorithm flow of the interactive differential privacy mechanism. This paper establishes a network evolution game model based on the social network structure and models the evolution process of users' privacy setting behaviors in online social networks. The experimental results show that the cost-benefit ratio of adopting a privacy mechanism has an important impact on the deployment and implementation of the privacy mechanism in social networks. The experimental results also show that the model proposed in this paper can effectively portray user privacy protection behaviors in social networks.

## 2. Related Work

Node attribute values are divided into identification information, quasi-identification information, and sensitive information [7]. Identification information is an attribute that can explicitly indicate an individual's identity, such as name and ID number; quasi-identification information can implicitly indicate an individual's identity, such as age and gender. Generally, a combination of multiple quasi-identification information is required to indicate an individual's identity. Sensitive information is information that needs to be protected, such as personal income and personal medical conditions. If the attacker knows the node attribute value he owns and matches the public social network data, it is possible to identify the true identity of the node and then obtain the user's sensitive information. At present, the attacker mainly matches and recognizes the nodes in the social network based on the background knowledge of some attack targets, so as to accurately or with a certain probability to identify the location of the attack target in the social network [8]. The process of the attacker matching and identifying the location of the attack target based on background knowledge is called node reidentification. For example, the nodes in the social network can be filtered according to the attribute values of multiple nodes, thereby further reducing the attack range of the attacker and increasing the attack hit rate [9].

Relevant scholars discussed how to implement the node  $k$ -anonymity model in social networks where each node has attribute information [10]. Gender and other generalization operations are performed, and  $k$  nodes with the same

quasi-identity attributes are divided into the same cluster after generalization, so that the attacker's hit rate of attacks based on quasi-identification information is reduced to  $1/k$ . Related scholars apply the  $(k, l)$  model of the database to the field of social network privacy protection [11]. It requires that on the basis of the  $k$ -anonymity model, there must be at least  $l$  users with different information in the cluster to ensure that the network resists the  $k$ -anonymity model attack. Related scholars proposed a  $k$ -degree anonymous model for node degree, which requires the number of nodes with the same degree attribute in the network to be greater than or equal to  $k$  [12]. At the same time, the  $k$ -degree anonymity requirement is achieved by adding extra edges. Researchers propose a  $k$ -neighborhood model to resist neighborhood attacks [13]. The model requires that each node has at least  $k - 1$  nodes with the same neighborhood structure and uses methods such as adding pseudo edges in the specific implementation.

Data perturbation mainly randomizes and modifies the original social network graph, so that the attacker cannot identify the target node based on the background knowledge he has mastered. At present, the general methods for realizing data disturbance include randomly adding pseudo edges or pseudo nodes, deleting nodes or edges, and modifying the attributes of nodes or edges. The edge weight represents the strength of the relationship between two nodes. Relevant scholars proposed the edge weight protection technology using greedy strategy in weighted undirected graphs to modify the weights of key edges under the premise of ensuring that the shortest path does not change, maintain the overall structure of the network before and after anonymity, and minimize the amount of information loss [14]. However, this article only considers the edge weight information and cannot resist other link attacks. Related scholars have evaluated Twitter's privacy policy and proposed the Hummingbird structure [15]. Hummingbird has made some changes on Twitter, which can protect the content of tweets and prevent hashtags from being obtained by centralized servers. However, these studies have always been based on the complete trust of centralized servers. In fact, centralized servers also have certain security risks and may leak user privacy. Although some studies encrypt and protect data on the server side, it still cannot prevent OSN service providers from monitoring user interactions, censoring or deleting user data, or even controlling who can establish social relationships with a social circle. Distributed structure does not rely on centralized servers, but it also faces a series of challenges. For example, encryption algorithms can ensure confidentiality and integrity while ensuring that they can provide high efficiency like centralized server structures.

Related scholars have proposed a multilevel security method [16]. In this method, trust is just a parameter used to determine the security level of a visitor or resource owner. Semidistributed autonomous access control was later proposed. It is an execution mechanism used to control information sharing in online social networks. This mechanism can also standardize the access rules of network resources. This is mainly achieved through different levels of authority between users. These levels are based on the type of social

relationship, the depth of the social relationship, and the security level. Researchers have proposed an online social network access control mechanism based on Web semantic technology, which is scalable and fine-grained access control [17]. Its main idea is to encode information related to social networks through ontology encoding technology. These studies are based on traditional access control methods, but these researchers have overlooked a problem: traditional access control may also leak the privacy of users' social attributes. Later, a user social attribute privacy protection scheme based on node splitting was proposed [18]. It increased the anonymity of the original node by assigning the attribute links and social links of the original node to the new node, thereby protecting the user's sensitive attributes from disclosure. At the same time, it also splits the sensitive social attributes of users according to the different degrees of influence of the social network structure on the distribution of social attributes and the correlation between social attributes [19, 20].

### 3. Social Network Big Data Analysis Platform

*3.1. Overall Design of Social Network Big Data Analysis.* For the same social network data, algorithms in different fields can be used for research from many angles. Increasing the utilization rate of data can also speed up upper-level research work through the unified interface of the platform and reduce the code strength in data acquisition and processing. From a functional point of view, the design and implementation of such a platform need to meet the following requirements:

- (1) Because users in social networks generate a large amount of data in real time, these data on the one hand supplement the user's historical data, and on the other hand, it completes the data of the entire relationship network. Therefore, incremental data acquisition capabilities for social networks are indispensable for the platform
- (2) When mass data is incorporated into a platform for unified management, these data usually need to have a fixed format or mode to facilitate the processing of upper-level applications. At the same time, in the context of big data, it is necessary to prevent the loss of data as much as possible. When the loss inevitably occurs, the platform also needs to be able to provide enough copies for data recovery, so an independent module is needed in the platform
- (3) It can provide fast calculation and algorithm expansion capabilities for massive data. In the massive data scenario, the ability to process data is a key factor in determining the availability of the platform, and whether the existing algorithms in the platform can be simply and efficiently extended is an important criterion for measuring the scalability of the platform. Therefore, the platform, as the manager of the entire cluster computing resources, needs to provide

fast computing services and algorithm extension interfaces for massive data

- (4) In the platform, not only need to understand the operating status of each component of the platform but also need to be able to easily view the progress and results of data analysis and algorithm operation. Therefore, visually displaying the status of all platforms through a unified interface not only facilitates the management and control of the platform but also increases the encapsulation of system modules so that users do not need to know the details of the platform

The social network big data analysis platform is built with Spark as the core and includes a one-stop platform for data acquisition, data processing, data mining, and data visualization. It has good openness, scalability, and versatility. The architecture of the system is shown in Figure 1, which mainly includes four modules: data capture, data preprocessing and storage, data mining and analysis, and data visualization.

*3.2. Distributed Social Network Crawler Workflow.* The reason why the crawler system adopts a distributed design is to improve the crawling efficiency of the crawler, but it also increases the difficulty of the system's task allocation, message communication, and error recovery. The distributed crawler system is functionally divided into a Master node and a Slave node. The main task of the Master node is to manage the status of the entire crawler system and schedule tasks, and the Slave node is responsible for crawling and parsing the assigned web links.

When the crawler successfully completes a crawling link, it will parse out the Weibo data or relational data saved in the current link. At this time, the crawler will send these data in the form of a message to the data collection component that exists on the Master side. If all the crawl links of a user are completed, a crawl task is completed. At this time, the data collection component will persist all the user data to disk and extract a user ID to be crawled from the user crawl queue.

When a Slave completes the registration process, the Master will extract a user ID to be crawled from the user crawling queue and create a crawling task. When this task is executed, the number of Weibo pages and the number of following pages of the user currently to be crawled will be spliced into URLs and placed in the crawl link queue. If the crawler sends a crawling task request, it will judge whether to send back the crawling task or add the crawler to the waiting queue according to whether the current crawling queue is empty. If the crawler sends the crawling exception information, the crawling link assigned to the crawler will be put into the crawling link queue again while waiting for the crawler to recover from the exception. If the crawler sends the crawling information, it will judge whether to send back the crawling task or add the crawler to the waiting queue according to whether the current crawling queue is empty. The crawler crawling process is shown in Figure 2.

*3.3. Data Preprocessing and Analysis Module.* HDFS is an open-source distributed file system implemented in Hadoop.

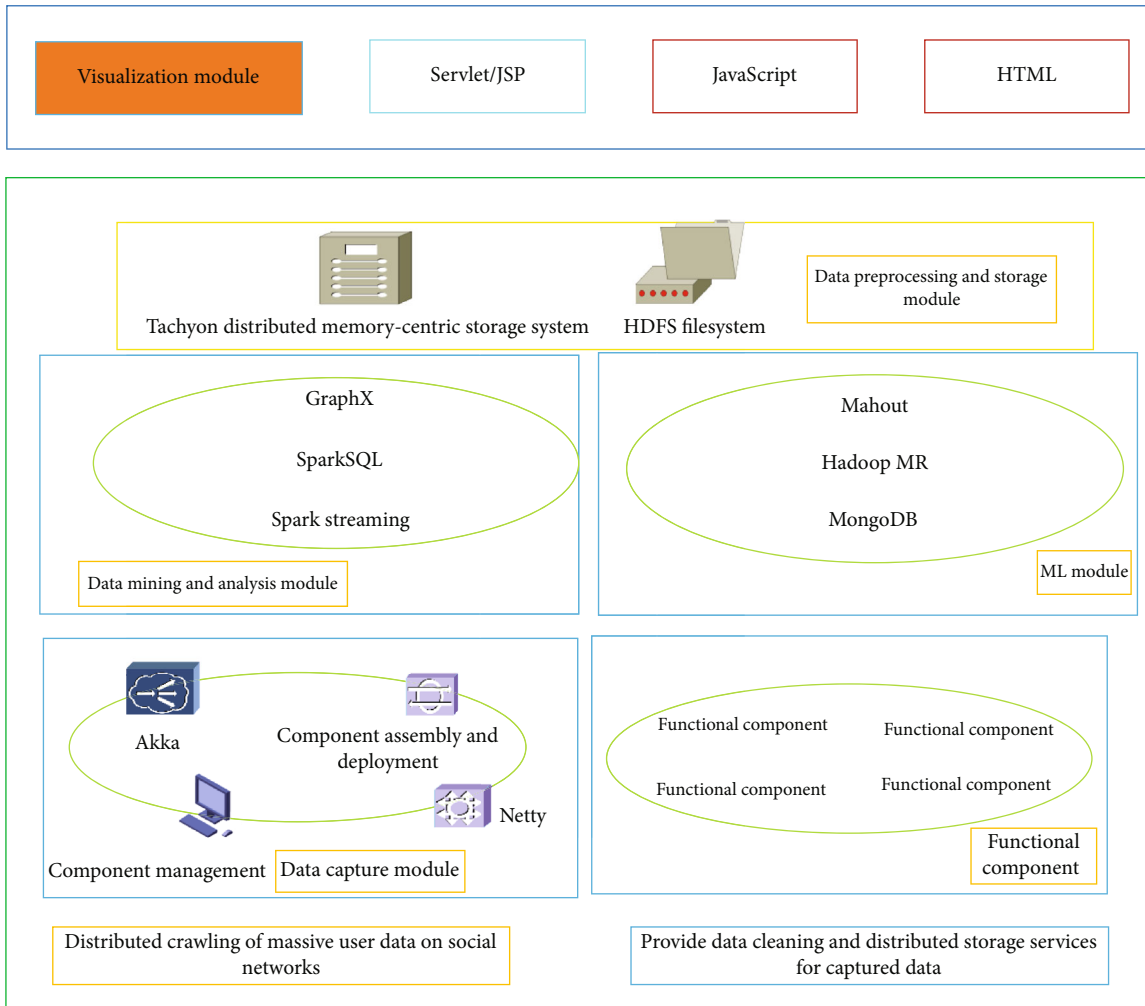


FIGURE 1: Social network big data analysis platform architecture.

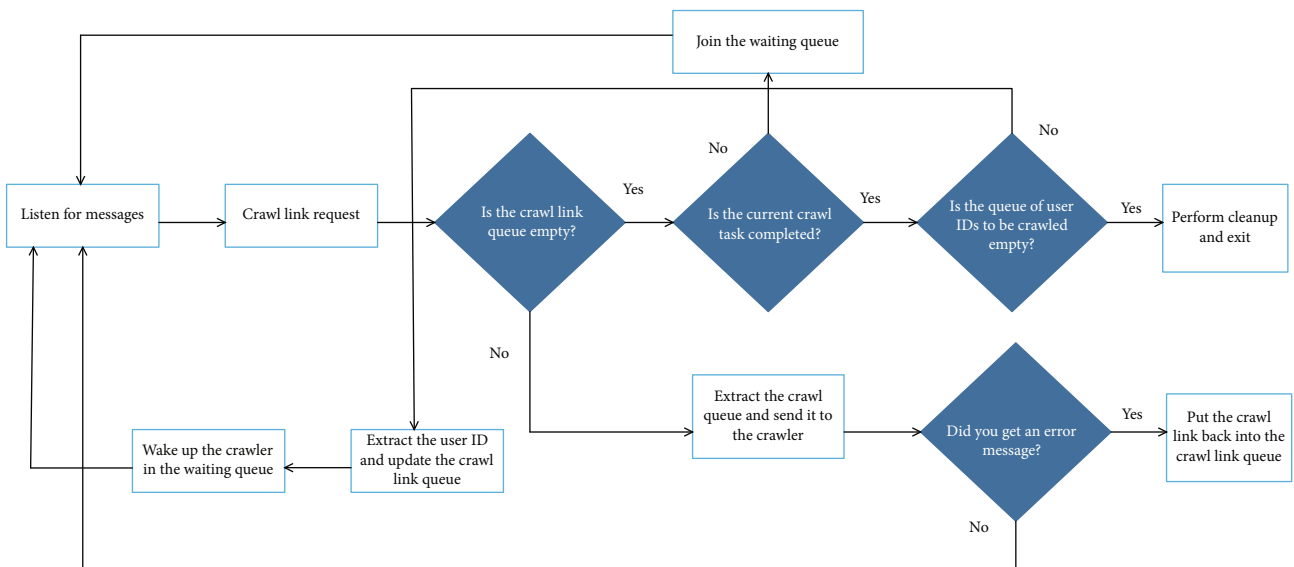


FIGURE 2: The crawler crawling process.

It can run on a cluster of a large number of cheap disks and provide reliable data slice storage and management, while ensuring faster data read and write speeds. The data preprocessing and storage module is basically the same as HDFS in function, so this module is mainly realized by building HDFS.

In order to ensure the versatility of the data analysis platform, so that the large amount of microblog data captured can serve more data mining and machine learning algorithms, the data preprocessing and storage module does not specify special features when storing the large amount of microblog data. Before the data is merged, the data needs to be formatted. The main work is to regularize the time field in Weibo and complete the default field to make it meet the structure of the database table. This ensures that upper-level applications can directly query text data through SQL (Spark SQL or Hive on Spark) when manipulating data.

The core component of the social network big data analysis platform is a data mining and analysis module built with Spark as the core. The purpose of this module is to use Spark's fast distributed computing capabilities and the MLlib machine learning components and GraphX graph computing components already provided by Spark. The machine learning and data mining algorithms implemented by the users themselves process the massive microblog data in the platform data storage module to complete the social network data mining and analysis tasks in the big data environment. All data mining and analysis tasks in this module will eventually be submitted to Spark to run in the form of Spark jobs. At present, there are two ways to run Spark: one is interactive operation in spark-shell, and the other is submitted through the Jar package mode to run offline. At the same time, Spark can also be used as a data source to read data from the outside using SQL statements through the JDBC interface.

When the social network big data analysis platform in this paper performs data processing and analysis tasks, Spark-JobServer will submit the Jar package to the deployed Spark cluster for execution. At the same time, in order to make the front-end more convenient to display the data, when the data visualization module requests the data in the platform, the data mining and analysis module will be used as the data source interface to complete the data query service through Hive on Spark.

From the user's point of view, there is no difference whether Hive is based on Spark or Hadoop, and data operations can be performed by connecting to the JDBC interface through the same statement. In this paper, the social network big data platform needs to query and access the data of the data preprocessing and storage module when data visualization. These operations will be implemented through the JDBC interface of Hive on Spark in the SQL interaction mode of the traditional relational database. The workflow of using Hive on Spark is shown in Figure 3.

## 4. Improved Differential Privacy Algorithm

*4.1. Principle Analysis of Differential Privacy.* Differential privacy initially achieved good results in the application of database statistical information. The differential privacy method

is developed based on the concept of "neighbors" data sets. The so-called "neighbor" data set concept refers to a data set that is different from the original data set, and the difference is only one record. This concept gives the most stringent definition of the differential privacy method, ensuring that the differential privacy method can resist attackers to the greatest extent and prevent the leakage of private data. The advantages of differential privacy make it highly regarded.

*4.1.1. Problem Description.* In order to describe the privacy protection problem more vividly, this article uses a connected undirected graph  $G$  to represent a social network, where each node of the graph represents an object in the social network, and each undirected edge of the graph represents the relationship between two objects. The mathematical notation is as follows:

In the graph  $G(V, E)$ ,  $V$  is the set of all nodes, and  $E$  is the set of all edges; the node set  $V$  is

$$V = \{v_i | i = 0, 1, 2, 3, \dots, n - 1\}, \quad (1)$$

where  $n$  is the number of nodes. The edge set is

$$E = [(v_i, v_j) | i, j = 0, 1, 2, 3, \dots, m - 1, i \neq j]. \quad (2)$$

Let  $q$  denote a query function of the graph and  $q(G)$  denote the result of the query function  $q$  acting on the graph  $G$ .

As we all know, when users use the query function  $q$  to obtain information from the social network graph  $G$ , the real result should be  $q(G)$ . In order to protect the private data on social networks, the real results cannot be returned to the user. This article needs to add some noise with a specific distribution on the basis of  $q(G)$ , so that the real information will not be leaked, and there is not much discrepancy between the real statistical results and does not affect the normal needs of users. This is the most basic principle of this article to achieve privacy protection.

*4.1.2. Definition and Implementation of Differential Privacy.* Traditional privacy protection methods often rely on the attacker's knowledge background, which causes them to have different flaws, which are only suitable for specific environments, and cannot achieve satisfactory results when acting on complex social networks. The proposed differential privacy protection method model perfectly solves the lack of privacy protection in social networks. It defines an extremely strict attack model, which realizes the protection of privacy by adding noise to the original data, the conversion of the original data, or the statistical results. Even if the attacker already knows all other data except the target data, the differential privacy mechanism can still achieve a good protection effect, ensuring that the target data will not be leaked.

$D_1$  and  $D_2$  are two data sets, and only one data is different between them.  $q$  represents a random function,  $\Pr[A]$  represents the probability of data being leaked, and  $S$  is a subset of all values of the function  $q$ , if the random function  $q$  satisfies

$$\Pr[q(D_1) \in S] < e^\epsilon \cdot \Pr[q(D_2) \in S]. \quad (3)$$

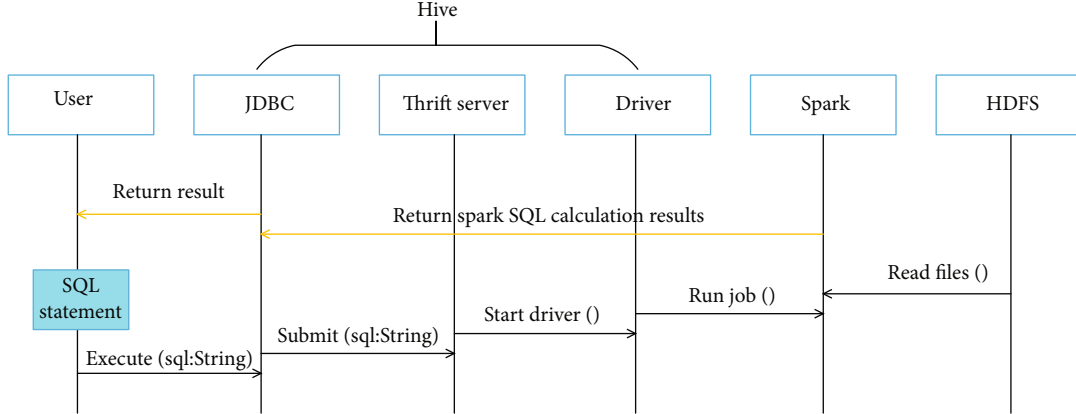


FIGURE 3: Hive on Spark workflow.

Then, it is said that  $q$  provides  $\epsilon$  differential privacy protection. There are differences between data sets  $D_1$  and  $D_2$ , and there is only one piece of data. This article refers to the original set and neighbor set, respectively. It can be seen that the realization of differential privacy is closely related to  $D_1$  and  $D_2$ . After research, it is found that differential privacy can be achieved by adding Laplacian distributed noise.

We call the distribution of the probability density function the Laplace distribution. The parameter  $\mu$  represents the position, and the parameter  $b > 0$  represents the scale.

$$f(x|u, b) = \frac{1}{2} \cdot b \cdot \begin{cases} \exp\left(\frac{u-x}{b}\right) & x \geq u, \\ \exp\left(\frac{x-u}{b}\right) & x < u. \end{cases} \quad (4)$$

In the Laplace distribution, this paper defaults  $\mu = 0$ , so the only uncertain parameter is  $b$ . And  $b$  is set based on the difference between the original set and its neighbor set.

For the query function  $q$ , this paper calls the difference between the original set and its neighbor set the sensitivity. Its sensitivity is defined as

$$\Delta q = \max [D_1 \quad D_2 \quad q(D_1) - q(D_2)]. \quad (5)$$

So far, the basic concepts related to differential privacy have been introduced. Its working mechanism can be understood as follows: a data set  $D$ , a query function  $q$ , and privacy control parameters  $\epsilon$  are input. Based on Laplace's noise, the probability distribution is

$$p(x|\lambda) = \frac{1}{2} \cdot \frac{e^{-x/\lambda}}{\lambda}. \quad (6)$$

Among them,  $\lambda$  is jointly determined by the sensitivity  $\Delta q$  and the control parameter  $\epsilon$ , and the sensitivity  $\Delta q$  is related to the data set  $D$  and its neighbor set. This working mechanism of differential privacy ensures that the attacker cannot accurately estimate the true "existence" of the target record even if he knows all the records except the target record and prevents the leakage of private data to the greatest extent.

**4.1.3. Analysis of the Effectiveness of Differential Privacy.** Through the analysis of the working mechanism of differential privacy, this article knows that the Laplacian noise mechanism guarantees the correctness of differential privacy. In the following, the subject thoroughly demonstrates the feasibility of Laplace noise and the effectiveness of the differential privacy protection method.

For the data set  $D$  and the query function  $q$  acting on  $D$ , the mechanism  $A$  is

$$A(D) = q(D - 1) + \text{Laplace}\left(\frac{\Delta q}{\epsilon}\right). \quad (7)$$

The Laplacian noise mechanism satisfies the protection conditions of  $\epsilon$  differential privacy and ensures the effectiveness of differential privacy. The goal of differential privacy is to maximize query accuracy and minimize the risk of privacy leakage. The noise of the Laplacian distribution makes the user or the attacker get a similar output for any result of the query function, regardless of whether the data set contains or does not contain the target record.

**4.2. Improved Differential Privacy.** The unique advantages of differential privacy make it attract strong attention as soon as it is proposed. Although the traditional differential privacy protection method can achieve good protection effects in the application of database statistical information, the complexity and data relevance of the social network itself hinder the application of the traditional differential privacy protection method. In response to this, this paper proposes an interactive differential privacy mechanism, so that the differential privacy protection mechanism can be better applied to social networks and better protect people's private data on social networks.

**4.2.1. Problem Description.** The key issue of differential privacy protection in social networks is to compare the original graph and its "neighbor graph" to calculate the sensitivity and adjust the parameter  $\epsilon$  to control privacy leakage. The traditional differential privacy protection model can get good results for individual users' independent queries (that is, there is no relationship between each query of the user),

but when the queries are closely related, it is difficult to avoid the leakage of private data. There are two main problems encountered in the application of differential privacy on social networks: First, on a graph such as a social network, the points or edges are all related to each other. When changing a point or an edge, the impact of changing a point or an edge is not only the independent point or this edge but also the point near this point or the edge near this edge and even the entire network. Second, under interactive conditions, there may be a relationship dependency between each query of the user. The current query result will depend on all previous query results. At this time, the sensitivity calculation needs to consider the current relationship between the results of each query while considering the impact of the query.

Summarizing the problems of traditional differential privacy applications in social networks, this article found a common phenomenon, and the most important point is that when comparing the original graph and its “neighbor graph” on social networks to calculate the sensitivity, it is not only necessary to calculate the current query situation, but it also needs to calculate the situation related to the current query. In this paper, this relationship factor between each other is called the interaction factor, and it is represented by  $t$ .

**4.2.2. Definition and Implementation of Interactive Differential Privacy.** Interactive differential privacy is implemented on the basis of traditional differential privacy. The main application object is social networks. The goal is to use social networks in a complex environment where “data” is closely related. Even if the attacker already knows all the other data except the target data, it can still achieve a good protection effect and ensure that the target data will not be leaked.

This paper still uses connected undirected graph  $G$  to represent a social network, and  $Q$  represents a query function for “social network.” Assuming that the user is currently querying the social network for the  $t + 1$ th time, the real result should be  $Q_t + 1(G)$ . Under the action of the interaction factor  $t$ , the result of the previous  $t$  times or  $t + 1$  times the result of the query will be affected by  $t$ . The results are  $Q_1(G), Q_2(G), \dots, Q_t(G)$ ; then,  $Q_t + 1(G)$  will depend on  $Q_1(G), Q_2(G), \dots, Q_t(G)$ . The interactive differential privacy protection mechanism not only needs to calculate the influence of  $Q_t + 1(G)$  but also needs to take into account all the influences including  $Q_1(G), Q_2(G), \dots, Q_t(G)$ .

$G$  and  $G'$  are two social network graphs,  $Q$  represents a random function,  $\Pr[A]$  represents the probability of data  $A$  being leaked, and  $S$  is a subset of all values of function  $Q$ . Then, as long as the following formula is satisfied, the query function  $Q$  realizes the differential privacy protection of  $\epsilon$ .

$$\Pr \{ [Q_{t+1}(G) | Q_1(G), \dots, Q_t(G)] \in S \} \leq \exp(\epsilon) \cdot \Pr \{ [Q_{t+1}(G') | Q_1(G'), \dots, Q_t(G')] \in S \}. \quad (8)$$

Interactive differential privacy still uses Laplacian distributed noise. Through the new sensitivity calculation formula

and constant adjustment of the parameter  $\epsilon$ , this article can obtain the Laplacian distribution of the noise that needs to be added and then achieve an interactive differential privacy method. Interactive differential privacy applications on social networks can also maximize query accuracy and minimize the risk of privacy leakage. In fact, although this article defines the dependency relationship between each query of the user, the query function is different for different practical scenarios, which causes the relationship between the query results to be different. Therefore, the interactive differential privacy protection model needs to refine the relationship between query results based on practical indicators and continuously adjust to achieve better protection effects.

**4.2.3. Analysis of the Effectiveness of Interactive Differential Privacy.** By adding Laplace distributed noise, this paper can still ensure the correctness of interactive differential privacy. For the data set  $D$  and the query function  $Q$  acting on  $D$ , the mechanism  $A$  is

$$A(D) = [Q(D)_t, Q(D)_{t-1}, \dots, Q(D)_1] - \text{Laplace}\left(\frac{\Delta Q}{\epsilon}\right). \quad (9)$$

This guarantees  $t * \epsilon$  differential privacy. When the interaction factor is  $t$ , and each query  $Q$  satisfies the  $\epsilon/t$  differential privacy, the mechanism will provide the guarantee of  $\epsilon$  differential privacy.

$$A(D) = [Q(D-1)_t, Q(D-1)_{t-1}, \dots, Q(D-1)_1] - \text{Laplace}\left(\frac{t \cdot \Delta Q}{\epsilon}\right). \quad (10)$$

Therefore, the interactive differential privacy mechanism proposed in this paper can also achieve  $\epsilon$  differential privacy by adding Laplacian distributed noise, maximize query accuracy, and minimize privacy leakage, while solving the complexity and data relevance of social networks.

**4.2.4. Algorithm Flow of Interactive Differential Privacy Mechanism.** The interactive differential privacy mechanism is implemented by adding Laplacian distributed noise. Unlike traditional differential privacy, in the sensitivity calculation process of interactive differential privacy, in addition to considering the impact of the current query, this article also needs to consider the interaction factor. The algorithm of the interactive differential privacy mechanism mainly includes two parts: the sensitivity solving process and the noise adding process. The flow of the interactive differential privacy algorithm is shown in Figure 4.

## 5. Experimental Analysis

**5.1. Experimental Design and Implementation.** This experiment will simulate the evolution process of the proportion of users who take privacy protection behaviors in the network on four sets of data sets and get the final evolution result by setting different costs and benefits. We compare the evolution result with the theoretical derivation to prove the correctness of the model. Three sets of data are real social

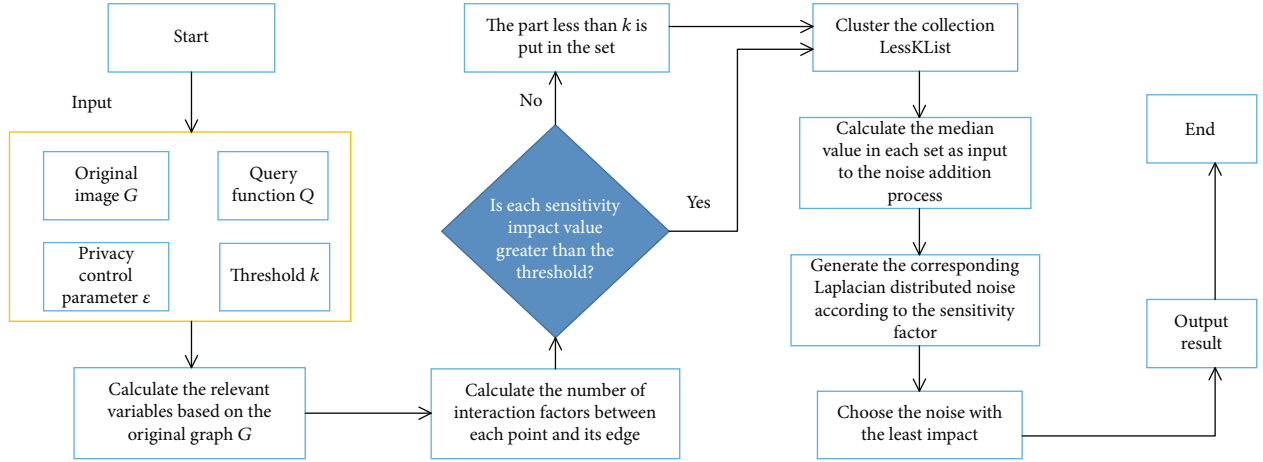


FIGURE 4: Interactive differential privacy algorithm flow.

network structure data, and the source is the Stanford large-scale network data collection SNAP project. The remaining set of data is man-made data, and the structure is a regular graph, that is, each point has the same degree. The data characteristics are shown in Figure 5.

**5.2. Analysis of Simulation Results.** In real social networks, most of the personal information is still in the hands of users, and connected friends only have a small part of the associated data, so the value of the correlation factor  $\alpha$  will not be very large, so we set  $\alpha$  to 0.1. Since evolutionary games occur under weak selection conditions, we set  $w$  to 0.001.

The evolution process of Ego-Facebook data set users' privacy protection behaviors under different benefit-cost ratios is shown in Figure 6. The proportion of users who initially adopt the privacy protection mechanism is set to  $p_0 = 0.4$ . Since the evolution process is relatively slow, in order to clearly show the data, the data points are average values. Although the abscissa of each data set is from 0 to 10000, each data point is the average of the previous  $N$  results. The number of updates is  $10000 \times N$ . Where  $c = 1$ ,  $b$  changes to achieve different benefit-cost ratios. According to the network structure and parameter settings, the theoretical result is that when  $b/c > 0.99$ , the proportion of users who adopt the privacy protection mechanism is finally 1. When  $b/c < 0.99$ , no one will finally adopt the privacy under the benefit-cost ratio protection mechanism. The simulation results show that the model can effectively evolve the dynamic evolution of user privacy protection behavior in social networks and can help social network managers design effective security services based on factors such as network structure, benefits, and costs. It can not only improve the security of the network but also reasonably calculate its own development costs.

When a social platform launches a new privacy protection setting or a user uses a certain privacy protection setting of the platform for the first time, there is no cost-benefit ratio that can be referred to in the past for strategy learning. At this time, users will decide whether to adopt it according to their privacy habits. Therefore, in the early stages of evolution, social networks had a proportion of initial users who adopted

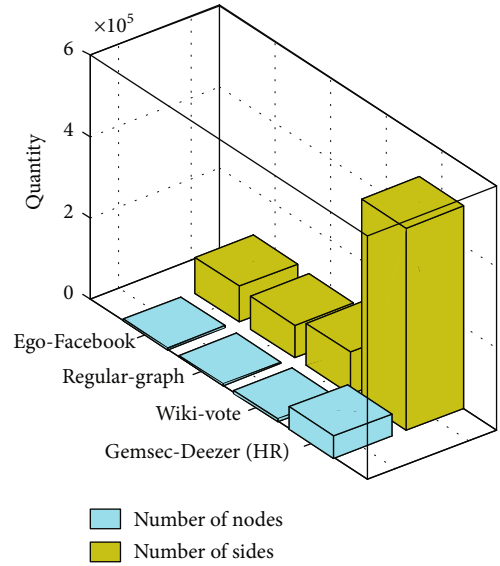


FIGURE 5: Simulation data set of user privacy behavior evolution process.

privacy protection settings. In order to verify the correctness, Figure 7 shows the simulation results of setting four different initial values  $p_0 = \{0.3, 0.4, 0.5, 0.6\}$  on the Ego-Facebook data set. Among them, the cost-benefit ratio in Figure 7(a) is  $c = 1.2$ , and the cost-benefit ratio in Figure 7(b) is  $c = 0.8$ .

**5.3. Comparison of Simulation Results of Each Data Set.** In order to illustrate the correctness of the model, this experiment conducted simulation experiments on four data sets with different structures. Figures 8–10 are the evolution results of regular-graph, Wiki-vote, and gemsec-Deezer (HR) data sets of user privacy behavior. From the experimental results, it can be seen that the curve of  $b/c > 1$  finally converges to 1, and the curve of  $b/c < 1$  finally converges to 0. When the average network degree approaches infinity, the critical value approaches 1. The experimental results prove that the model is feasible. At the same time, it is found that as the number of nodes in the network increases, the evolution process becomes slower and slower. The data set used

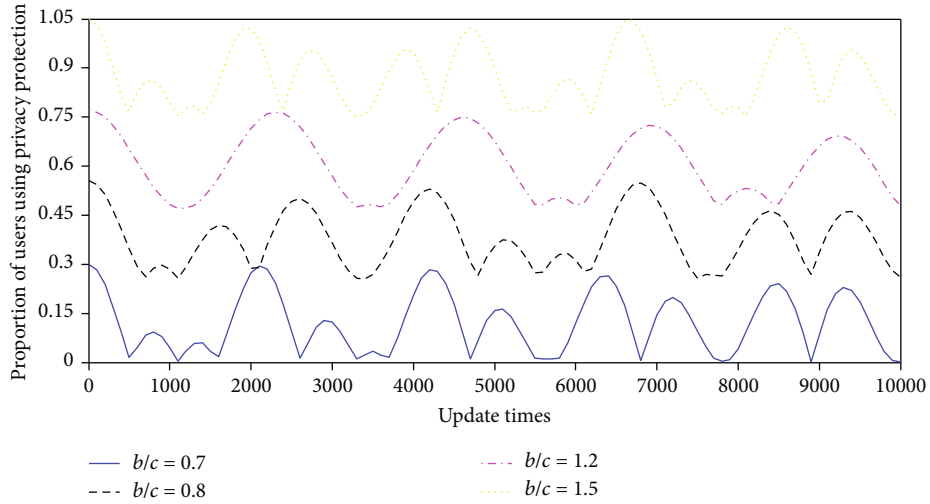
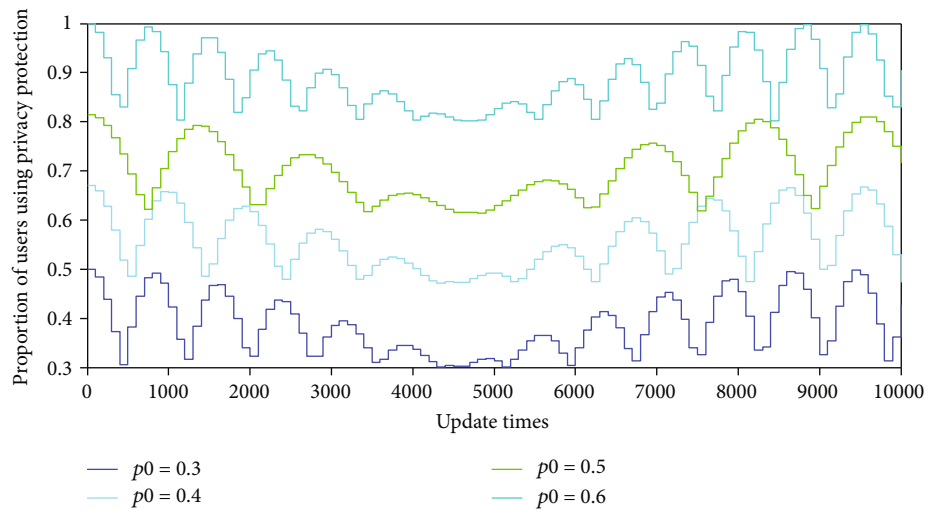
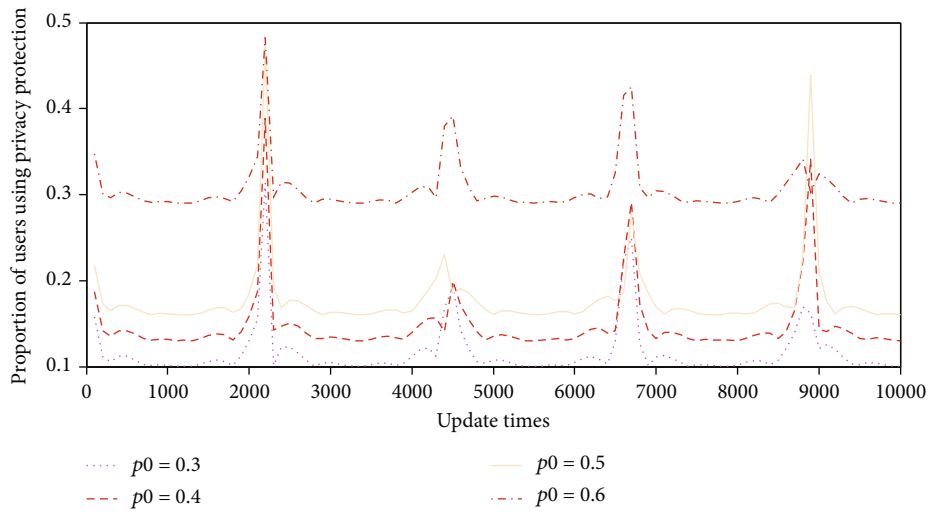


FIGURE 6: Ego-Facebook data set privacy protection behavior evolution simulation results.



(a) The cost-benefit ratio is  $c = 1.2$



(b) The cost-benefit ratio is  $c = 0.8$

FIGURE 7: Evolution results of different initial user proportions.



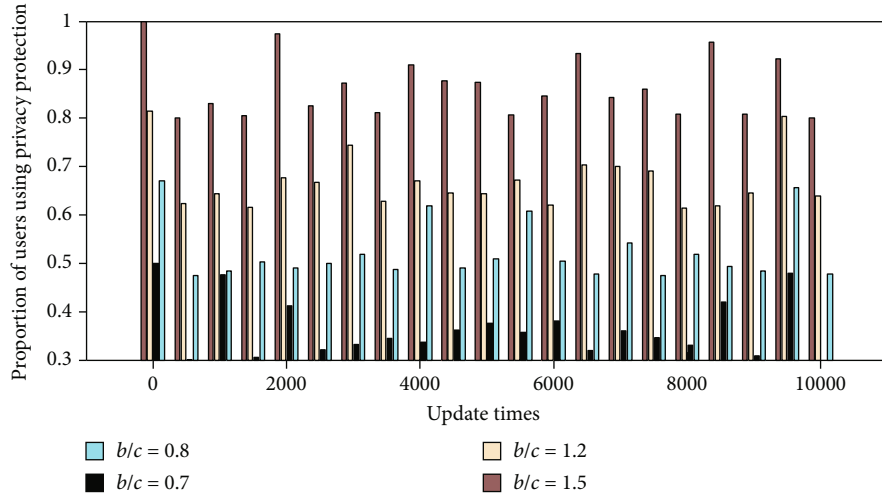


FIGURE 8: The evolution result of the privacy protection behavior of the regular-graph data set.

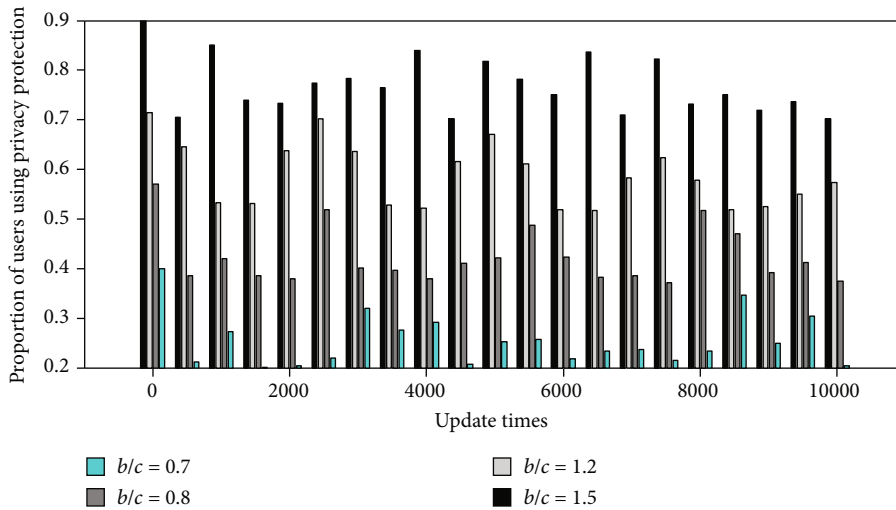


FIGURE 9: Wiki-vote data set privacy protection behavior evolution result.

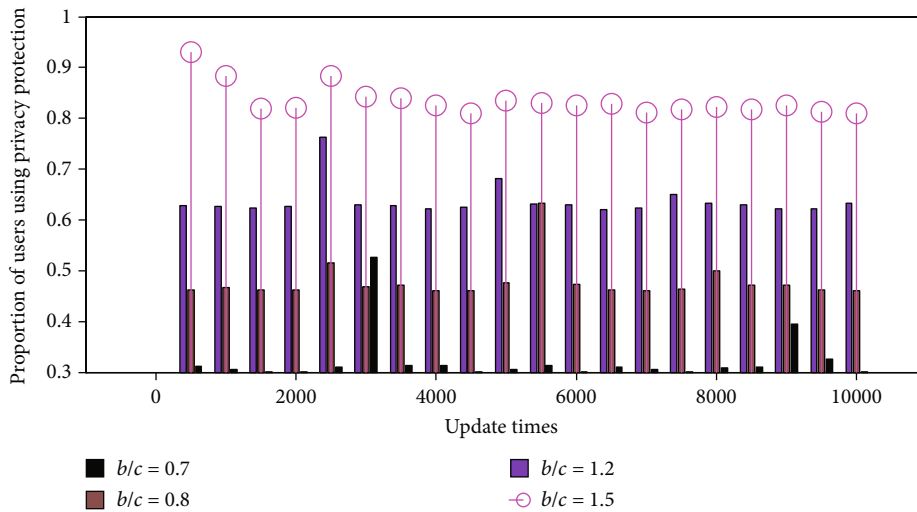


FIGURE 10: The evolution result of the privacy protection behavior of the gemsec-Deezer (HR) data set.

in Figure 8 is an artificial data set, and the number of nodes is the same as the Ego-Facebook data set, but the connection structure is different. The update time required for the regular structure to converge to the result is longer than the time required for the actual heterogeneous social network structure, indicating that the heterogeneity of the network structure can accelerate the curve convergence. It can be found that the evolution process varies with the network structure, but the evolution trend follows the theory derived from the model, which verifies the correctness of the model. Experimental results show that the model can effectively model the dynamic evolution process of user privacy protection behavior in social networks.

## 6. Conclusion

Aiming at the status quo that traditional differential privacy methods cannot be applied to interactive social networks, the subject research first designed a novel differential privacy protection mechanism based on interaction factors. Relying on this new mechanism, this paper can not only prevent the leakage of private data to the greatest extent as the traditional differential privacy method but also solve the problem of privacy leakage caused by the characteristics of the data relevance of the social network itself. This paper proposes a user security behavior game model based on network evolutionary game theory and defines the basic elements of the game model and the updated rules of the evolutionary game. The dynamic equation is derived based on the game model, which shows the evolution process of the benefit-cost ratio of the user's security behavior under the social network structure with the privacy protection provided by the platform. In order to verify the effectiveness of the model, simulation experiments were designed on artificial data sets and real data sets. Experimental results show that the model can effectively describe the evolution of user safety behavior. In the social attribute privacy protection scenario, although the method proposed in this paper solves the privacy problem of social attribute leakage that may occur during access control in social networks, there is also social attribute privacy leakage in other scenarios, such as users in data mining. For sensitive attribute leakage, how to protect user attribute privacy while ensuring data availability is the next problem to be solved.

## Data Availability

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

## Conflicts of Interest

We declare that there is no conflict of interest.

## References

- [1] Q. Wang, Y. Zhang, X. Lu, Z. Wang, Z. Qin, and K. Ren, "Real-time and spatio-temporal crowd-sourced social network data publishing with differential privacy," *IEEE Transactions on Dependable and Secure Computing*, vol. 15, no. 4, pp. 591–606, 2018.
- [2] R. Chen, B. C. Fung, P. S. Yu, and B. C. Desai, "Correlated network data publication via differential privacy," *VLDB Journal*, vol. 23, no. 4, pp. 653–676, 2014.
- [3] J. M. Such and N. Criado, "Resolving multi-party privacy conflicts in social media," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 7, pp. 1851–1863, 2016.
- [4] Q. Fang, J. Sang, C. Xu, and M. S. Hossain, "Relational user attribute inference in social media," *IEEE Transactions on Multimedia*, vol. 17, no. 7, pp. 1031–1044, 2015.
- [5] P. Wang, Z. Gao, X. Xu, Y. Zhou, H. Zhu, and K. Q. Zhu, "Automatic inference of movements from contact histories," *ACM SIGCOMM Computer Communication Review*, vol. 41, no. 4, pp. 386–387, 2010.
- [6] E. K. Wang, Y. Li, Y. Ye, S. M. Yiu, and L. C. K. Hui, "A dynamic trust framework for opportunistic mobile social networks," *IEEE Transactions on Network and Service Management*, vol. 15, no. 1, pp. 319–329, 2018.
- [7] N. Halko, P. G. Martinsson, and J. A. Tropp, "Finding structure with randomness: probabilistic algorithms for constructing approximate matrix decompositions," *SIAM Review*, vol. 53, no. 2, pp. 217–288, 2011.
- [8] K. Xu, Y. Guo, L. Guo, Y. Fang, and X. Li, "My privacy my decision: control of photo sharing on online social networks," *IEEE Transactions on Dependable and Secure Computing*, vol. 14, no. 2, pp. 199–210, 2017.
- [9] R. Heatherly, M. Kantarcioglu, and B. Thuraisingham, "Preventing private information inference attacks on social networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 8, pp. 1849–1862, 2013.
- [10] M. Vegni and V. Loscri, "A survey on vehicular social networks," *IEEE Communication Surveys and Tutorials*, vol. 17, no. 4, pp. 2397–2419, 2015.
- [11] S. Rathore, P. K. Sharma, V. Loia, Y.-S. Jeong, and J. H. Park, "Social network security: issues challenges threats and solutions," *Information Sciences*, vol. 421, pp. 43–69, 2017.
- [12] E. Palomar, Á. Galán, A. Alcaide, and L. González-Manzano, "Implementing a privacy-enhanced attribute-based credential system for online social networks with co-ownership management," *IET Information Security*, vol. 10, no. 2, pp. 60–68, 2016.
- [13] J. He, L. Cai, and X. Guan, "Preserving data-privacy with added noises: optimal estimation and privacy analysis," *IEEE Transactions on Information Theory*, vol. 64, no. 8, pp. 5677–5690, 2018.
- [14] N. Korula and S. Lattanzi, "An efficient reconciliation algorithm for social networks," *Proceedings of the VLDB Endowment*, vol. 7, no. 5, pp. 377–388, 2014.
- [15] B. Zhou and J. Pei, "The  $k$ -anonymity and  $l$ -diversity approaches for privacy preservation in social networks against neighborhood attacks," *Knowledge and Information Systems*, vol. 28, no. 1, pp. 47–77, 2011.
- [16] H. Hu, G.-J. Ahn, and J. Jorgensen, "Multiparty access control for online social networks: model and mechanisms," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 7, pp. 1614–1627, 2013.
- [17] B. Carminati, E. Ferrari, and M. Viviani, "Security and trust in online social networks," *Synthesis Lectures on Information Security, Privacy, and Trust*, vol. 4, no. 3, pp. 1–120, 2013.
- [18] Z. Zhang, Q. Gu, T. Yue, and S. Su, "Identifying the same person across two similar social networks in a unified way: globally and locally," *Information Sciences*, vol. 394, pp. 53–67, 2017.

- [19] J. M. Such and M. Rovatsos, "Privacy policy negotiation in social media," *ACM Transactions on Autonomous and Adaptive Systems*, vol. 11, no. 1, pp. 1–29, 2016.
- [20] H. Li, H. Zhu, S. Du, X. Liang, and X. Shen, "Privacy leakage of location sharing in mobile social networks: attacks and defense," *IEEE Transactions on Dependable and Secure Computing*, vol. 15, no. 4, pp. 646–660, 2018.

## Research Article

# A Novel Smart Depression Recognition Method Using Human-Computer Interaction System

Lijun Xu <sup>1</sup>, Jianjun Hou,<sup>1</sup> and Jun Gao<sup>2</sup>

<sup>1</sup>Institute of Art and Design, Nanjing Institute of Technology, Nanjing, Jiangsu 211167, China

<sup>2</sup>Siemens Ltd., China Jiangsu Branch Co., Ltd., Nanjing, Jiangsu 211100, China

Correspondence should be addressed to Lijun Xu; xulijun@njit.edu.cn

Received 11 April 2021; Revised 8 May 2021; Accepted 17 May 2021; Published 26 May 2021

Academic Editor: Shan Zhong

Copyright © 2021 Lijun Xu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In recent years, depression not only makes patients suffer from psychological pain such as self-blame but also has a high disability mortality rate. Early detection and diagnosis of depression and timely treatment of patients with different levels can improve the cure rate. Because there are quite a few potential depression patients who are not aware of their illness, some even suspect that they are sick but are unwilling to go to the hospital. In response to this situation, this research designed an intelligent depression recognition human-computer interaction system. The main contributions of this research are (1) the use of an audio depression regression model (DR AudioNet) based on a convolutional neural network (CNN) and a long-short-term memory network (LSTM) to identify the prevalence of depression patients. And it uses a multiscale audio differential normalization (MADN) feature extraction algorithm. The MADN feature describes the characteristics of nonpersonalized speech, and two network models are designed based on the MADN features of two adjacent segments of audio. Comparative experiments show that the method is effective in identifying depression. (2) Based on the research conclusion of the previous step, a human-computer interaction system is designed. After the user inputs his own voice, the final recognition result is output through the recognition of the network model used in this research. Visual operation is more convenient for users and has a practical application value.

## 1. Introduction

Depression is a hidden mental illness, and it also involves mental health problems. Symptoms may be manifested in emotional and emotional disorders and physical discomfort. Depression not only harms patients [1, 2] but also brings a heavy economic burden to patients' families and society. Depression can lead to increased expenditure on medicines, psychotherapy, rehabilitation, and other aspects, and it can also make patients inefficient or even unable to work. The report in Reference [3] pointed out that in 2002, the United States suffered as much as 44 billion U.S. dollars in economic losses due to depression causing workers to be unable to work normally or reducing work efficiency. Although there is no relevant data in China, the huge economic losses caused by depression can still be seen from the above figures.

Although depression is hugely harmful, it is a disease that can be effectively treated and improved. In clinical practice,

medication can promote the recovery of depression patients. In addition, psychotherapy and physical therapy can also achieve better results. However, a considerable part of the patient population has not been diagnosed in time. More than half of the people affected by depression at home and abroad have not received treatment. There are many reasons, such as insufficient precision in quantifying the prevalence of depression [4–6]. In addition, the diagnosis of depression is misdiagnosed. For example, depression and bipolar disorder are two relatively similar mood disorders. The depressive periods of depression and bipolar disorder are often difficult to distinguish [7]. Global Burden of Disease (GBD) conducted multiple epidemiological surveys for middle school students in 2016 [8]; the number of depressive symptoms in middle school students ranges from one-fifth to one-half.

The key to the treatment of depression is the preliminary diagnostic screening. If it is possible to quickly diagnose whether an individual is suffering from depression under rel-

actively safe conditions without too much privacy involved, it will greatly reduce the difficulty of clinical screening for depression and encourage patients to receive treatment as soon as possible. Voice is a noninvasive, clinically accessible information. At present, there have been a large number of studies on speech and depression [9, 10], which provides the possibility to explore speech as a tool for automated diagnosis of clinical depression. According to existing studies, the speech of patients with depression has the following characteristics: slower speaking rate, frequent pauses, long pauses [11, 12], reduced changes in voice characteristics [13], lack of circumflex and frustration, and dull voice [14]. Compared with normal individuals, individuals in the depression group have more pronounced breath sounds [15]. From the perspective of prosody characteristics, the changes in fundamental frequency (F0) of depression patients decrease, such as bandwidth, amplitude, and energy [16], which indicates that the changes in voice frequency of depression patients decrease. The spectral characteristics are also related to the degree of depression of the patient. Studies have found that the degree of change in the sound spectrum energy below 500 Hz and 500-1000 Hz is related to the severity of depression [17]. It can be seen that feature extraction of speech helps to better understand depression.

This research is based on speech data to recognize depression. In the early stage, the speech signal characteristics of the 2014AVEC dataset were extracted, and algorithmic modeling was used to identify the degree of depression of the subject. The model with relatively good experimental effect is selected to realize the human-computer interaction system in the later depression recognition. Finally, the experimental subject of research and application of speech-based depression recognition is completed. The work of this research includes the following:

- (1) The MADN method is used for speech feature extraction. Since the extracted conventional speech features include the sample's own personalized speaking features, this personalized speaking feature will affect the training of the depression recognition model, resulting in poor generalization of the trained model. Therefore, this study selects the MADN method for feature extraction, and the features extracted based on this method have better adaptability
- (2) The purpose of traditional depression recognition is to identify whether you have depression. The purpose of this study is not only to identify whether you have depression but also to determine the degree of depression. Only by identifying the specific degree of depression can the most appropriate treatment plan be given. The depression recognition model used in this study is DR AudioNet
- (3) Design a human-computer interaction system where users can input audio data online to identify depression. The system can help users understand their own depression status and can also assist doctors in initial examinations

## 2. Speech-Based Depression Recognition

In order to identify depression more conveniently and quickly, this paper designs a set of human-computer interaction system. First, the system will collect voice data. Secondly, the sample set is preprocessed and feature extracted at the front end. Again, machine learning algorithms [18, 19] and deep learning algorithms [20, 21] are used to model the samples in the backend. Finally, the trained model is used to obtain the recognition result of the test sample, thereby obtaining the final depression recognition result. The speech-based depression recognition process is shown in Figure 1.

The detailed process of identifying depression is as follows:

- (1) Perform preprocessing and feature extraction on the original data. The extracted features include Mel frequency cepstrum coefficients (MFCCs), zero-crossing rate, energy, and other speech features. These feature data will be used for model training
- (2) The front end is used for feature data collection (back-end auxiliary data processing). The data processing models mainly include deep learning algorithms such as CNN and deep convolutional neural network (DCNN), and you can also choose a suitable machine learning algorithm
- (3) The results identified by the back-end can be output to the visual display page according to the interaction requirements and displayed to doctors and patients. Doctors and patients have a preliminary understanding of the condition based on the output of the system

Although the research on the recognition of depression based on speech signals has achieved good research results at this stage, it is inevitable that there are still challenges in this research field.

Mainly reflected in the following aspects:

- (1) Effectiveness of voice feature selection: the classification of depression and nondepression based on speech signals is based on the premise that the speech features with significant differences between the two can be extracted. Different people have different characteristics in terms of tone, loudness, and energy of voice signals. The key point is to be able to find the characteristics that distinguish depressed individuals from nondepressed individuals in order to achieve better recognition results. However, relevant research results show that this effective voice feature has not been found yet. The most effective speech features on different speech datasets may be different
- (2) The accuracy of the results of depression recognition: due to different data characteristics, experimental algorithms, and hardware equipment conditions, recognition results with different accuracy will naturally be obtained. To improve the accuracy of the recognition results of depression by continuously improving

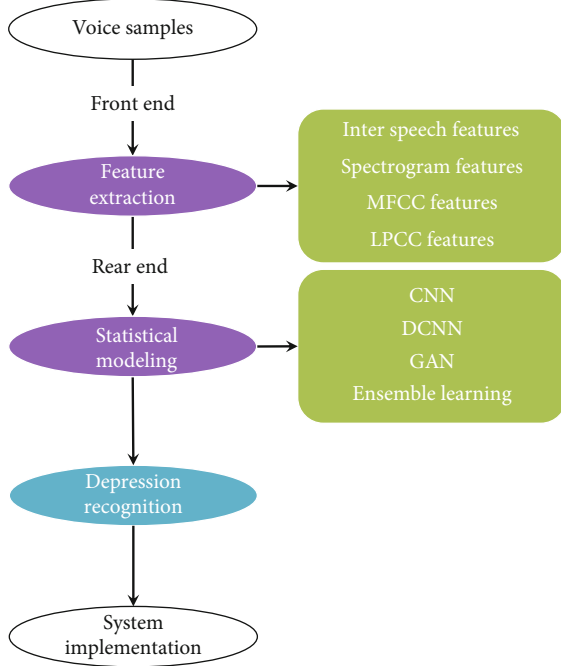


FIGURE 1: Depression recognition process.

the experimental conditions in terms of software and hardware is a problem to be solved in this research

- (3) The authority of the depression speech dataset: at the present research stage, there is no unified standard dataset as the speech source material for depression recognition. Some of the datasets used for the research are the voices of the subjects collected by the experimental researchers themselves, and some use the international open source voice datasets. The voice file collection process is also affected by the field environment, the denoising effect of the equipment, and the immediate state of the subjects. For this reason, seeking a unified standard for the integration of speech data related to depression is one of the challenges facing this field

### 3. Voice-Based Depression Recognition Method Used

**3.1. Depression Recognition Model Used.** The network model used in this study integrates CNN and LSTM. The type of data processed is voice data, and the output is whether it is a depression patient. The deep learning model used is DR AudioNet, and its network structure is shown in Figure 2.

As can be seen from Figure 1, the DR AudioNet network mainly consists of 2 convolutional layers, 1 pooling layer, 2 LSTM layers, and 1 fully connected layer. The input of the network is voice feature data. The construction process of speech feature data is as follows: extract MFCCs, zero-crossing rate, energy, and other features for each frame of speech in each speech, and select 60 consecutive frames of speech characteristics in each speech to construct a two-dimensional matrix. The X axis of the matrix is time, and

the Y axis is frequency information. The role of the convolutional layer is to extract the semantic information of Gao. The role of the pooling layer is to reduce the dimension of features. The two-dimensional matrix is converted into one-dimensional data after passing through the convolutional layer and the pooling layer. The LSTM layer is used to extract long-term dependency information. The fully connected layer is used to encode changes in speech on the X axis and give prediction results.

In order to make full use of speech information to improve the recognition rate, this paper uses an improved network model based on the DR AudioNet network. The improved network mainly contains 3 models, namely, M1, M2, and M3. These 3 models are all DR AudioNet networks. The execution process of the improved network is to first extract the feature V1 of the current speech segment. Feature V1 mainly includes MFCCs, short-term energy, short-term zero-crossing rate, and formant frequency. Input feature V1 into the M1 model. Considering that the personalized interference information mixed in each sample will affect the recognition effect of the model, the feature V2 of the previous segment of the current speech segment is used to train M2. Feature V2 is obtained using the MADN method. In order to further reduce the interference of the personalized information carried by the sample, the feature V3 of the last segment of the current speech segment is used to train M3. Feature V3 is obtained using the MADN method. The structure of the improved network is shown in Figure 3.

#### 3.2. Multiple Feature Extraction

**3.2.1. Mel Frequency Cepstrum Coefficients (MFCCs).** MFCCs are one of the most widely used and basic voice features. Equation (1) gives the conversion relationship between ordinary frequencies and MFCCs.

$$f_{\text{mel}} = 2595 \log \left( 1 + \frac{f_{\text{Hz}}}{700} \right), \quad (1)$$

where  $f_{\text{mel}}$  represents Mel frequency scale and  $f_{\text{Hz}}$  stands for normal frequency. Use  $P$  filters to calculate MFCCs. The center frequencies of these filters can be evenly arranged according to the Mel frequency, and the frequencies of the two bottom points of each filter are the center frequencies of the two adjacent filters. The output is  $L(p)$ ,  $p = 1, 2, \dots, P$  after filtering. Let  $l(p)$  be the lower limit frequency,  $c(p)$  be the center frequency, and  $h(p)$  be the upper limit frequency of the  $p$ -th triangle filter. Then, the relationship between them is as follows:

$$c(p) = h(p-1) = l(p+1). \quad (2)$$

The output of the filter is transformed as follows to obtain MFCCs.

$$C_n = \sum_{p=1}^P \log L(p) \cos \left( (p-0.5) \frac{h\pi}{P} \right) \quad h = 1, 2, \dots, H, \quad (3)$$

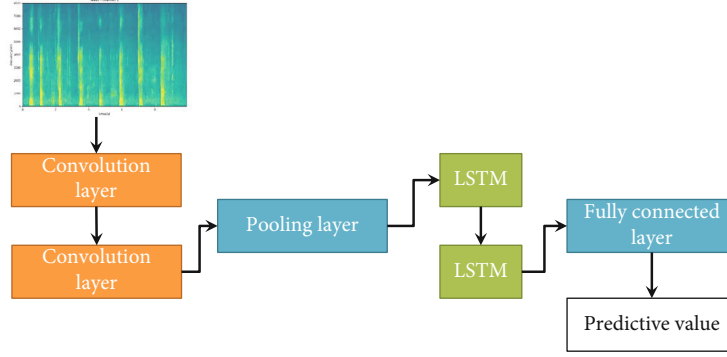


FIGURE 2: DR AudioNet network structure.

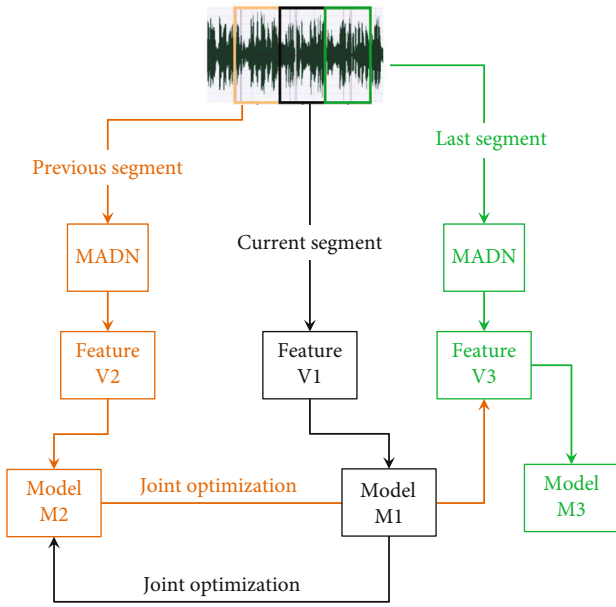


FIGURE 3: Structure diagram of the network used in this research.

where  $H$  represents the number of MFCCs,  $H \in \{12, 13, 14, 15, 16\}$ ,  $20 \leq P \leq 40$ .

**3.2.2. Resonant Peak, Energy, and Zero Crossing Rate.** Energy usually refers to the average energy of a frame of speech signal, which represents the changing trend of the signal. It is possible to analyze whether the voice signal contains sound through energy analysis. By analyzing the voice characteristics of depression patients, it can be observed that the voice characteristics of depression patients mostly include intermittent, unclear speech, and long pauses. Therefore, energy characteristics can be used to identify depression. The so-called zero crossing rate refers to the number of times the signal value passes through the zero value per second. The zero-crossing rate feature has been widely used in the fields of speech recognition and music information retrieval and is a key feature for classification of percussive sounds. The combination of this feature and energy feature is one of the most commonly used feature combinations in speech recognition.

**3.3. MADN Method.** In the actual scene, each person has a different tone color and voice size, some people have a high pitch, and some people have a low voice. The existence of this voice feature reduces the accuracy of depression recognition, and the personalized speech feature will weaken the generalization ability of the depression recognition model. This is because many features such as MFCCs extracted from each frame of audio include not only depression-related features but also static characteristics of voice personality. In response to this, this article uses the MADN algorithm to identify and remove personalized features of depression. The calculation process of the above MADN algorithm is as follows:

- (1) Read the original audio
- (2) Preprocess all audio
- (3) Extract features such as MFCCs, denoted by  $V(k, s)$ ,  $s$  is the number of frames of speech, and each frame contains  $k$  elements
- (4)  $D(k, s)$  is obtained by differential calculation through the feature  $V(k, s)$  of two adjacent audio segments, and the differential processing is used to eliminate the personalized information in the voice.  $D(k, s)$  represents the time sequence change of audio, and the distribution of its characteristic value is relatively stable. The calculation is described as follows:  $S$  is the total number of frames of speech

$$D(k, s) = V(k, s + 1) - V(k, s), \quad s = 1, 2, \dots, S - 1 \quad (4)$$

- (5) Normalize different scales for different features

$$F(k, s) = \frac{D(k, s) - D_{\min}(k, W_k : S_k)}{D_{\max}(k, W_k : S_k) - D_{\min}(k, W_k : S_k)}, \quad k = 1, 2, \dots, K \quad (5)$$

The values of  $S_k$  and  $W_k$  are different scales and sliding

windows, and the calculation formula is

$$W_k = \begin{cases} \max(0, k-5) & k = 1, 2, \dots, 12, \\ \max(0, k-10) & k = 13, 14, \\ \max(0, k-15) & k = 15, 16, 17, \end{cases} \quad (6)$$

$$S_k = \begin{cases} \min(60, k+5) & k = 1, 2, \dots, 12, \\ \min(60, k+10) & k = 13, 14, \\ \min(60, k+15) & k = 15, 16, 17 \end{cases} \quad (7)$$

(6) Output normalized features  $F(k, s)$  of various scales

## 4. Simulation Experiment and Analysis

**4.1. Experiment-Related Settings.** In order to analyze the superiority of the network used in identifying depression, the evaluation indicators used are Mean Absolute Error (MAE) [22] and Root Mean Square Error (RMSE) [22]. The experiment mainly conducts research from two aspects; one is the influence of different feature extraction methods on the results. The second is to use different regression algorithms. The feature extraction methods used include separate MFCCs and multifeature combination. Comparison algorithms are logistic regression (LR) [23], CNN [24], and DCNN [25]. The parameter settings of the network used in this research are shown in Table 1.

### 4.2. Experimental Results and Analysis

**4.2.1. Performance Comparison of 3 Models in the Network Used.** In order to verify whether the three models in the used network have the expected optimization function, this study conducted experiments on the 2014AVEC test set. Figure 4 gives the recognition results of the three models on the voice data in the network. In Figure 4, from M1 to M3, the values of RMSE and MAE are gradually decreasing, which shows that after the optimization of the previous model, the performance of the latter model has indeed been improved. The reason for the gradual improvement of the recognition performance of the three models is that on the basis of the DR AudioNet network; the upper-level model is optimized using the features of adjacent voices, thereby improving the recognition effect.

**4.2.2. Comparison of the Recognition Effect of Different Models on Depression.** After comparing the performance of the three models in the improved network, we will further explore the effects of other features and models in the recognition of depression. Through comparative experiments, to demonstrate the effectiveness of the feature V1 and DR AudioNet network used in this article, Table 2 shows the depression recognition effect of different features and different models.

From the experimental results of the LR, CNN, and DCNN models, it can be seen that although the V1 feature is a combination of multiple features, the experimental

results on the V1 feature are not better than the MFCCs alone. On DCNN, the recognition effect of the two features is not much different. On the AudioNet model, the results of the V1 feature are significantly better than the MFCCs alone. Furthermore, the network used in this paper recognizes the three characteristic data of V1, V2, and V3; the recognition result is obviously improved; and the performance advantage is obvious. Through comparative experiments, it can be concluded that the characteristics of the network used in this article V1 and the DR AudioNet network can effectively predict the degree of depression. When extracting features, different scales are used to normalize features, which effectively merges different features and retains audio depression features. At the same time, V2 and V3 are used to jointly optimize DR AudioNet, which effectively integrates the non-personalized depression characteristics of the MADN feature to the speaker.

**4.2.3. Comparison of Noise Immunity.** Noise immunity is a key indicator to measure the performance of an algorithm, because the actual data will inevitably be mixed with some noise data. These noisy data are bound to have a certain impact on the recognition results. An algorithm with good performance should have good noise immunity. In order to explore whether the algorithm in this paper is susceptible to noise, here, we add different amounts of Gaussian white noise to the dataset. During the experiment, this study added Gaussian white noise with a mean value of 0 and a variance of 0.01, 0.02, 0.03, 0.04, and 0.05 to the original dataset. Each model uses V1 features, and the evaluation indicators still use MAE and RMSE. The recognition results of each model on the noisy dataset are shown in Table 3.

The data in Table 3 shows that as the noise variance increases, the recognition performance of all models shows a downward trend. This is consistent with the theory. Among them, the performance decline trend of LR, CNN, and AudioNet is obvious. DCNN is not significantly affected by noise, and its noise immunity is the best overall. When the variance is less than 0.03, the network used is not sensitive to noise and has strong robustness. However, as the variance gradually increases, the performance of the network used begins to decline, and the downward trend is obvious, resulting in a lower performance than the DCNN model. In summary, the network used is suitable for processing micronoise data. Although the network used has certain antinoise performance, its antinoise performance needs to be further improved.

## 5. Design of Human-Computer Interaction System for Intelligent Recognition of Depression

In order to assist doctors in the preliminary diagnosis of patients with suspected depression, based on the above research results, this paper designs a human-computer interaction system for intelligent recognition of depression. The overall design requirements of the system are shown in Figure 5.



TABLE 1: Parameter settings.

Parameter	Value
Feature size	17 * 60
Batch size	32
The number of convolution kernels in the convolution layer	64
Convolution kernel size	3 * 1
The number of cells in the LSTM layer	128
Number of nodes in the first fully connected layer	128
The last fully connected layer	1

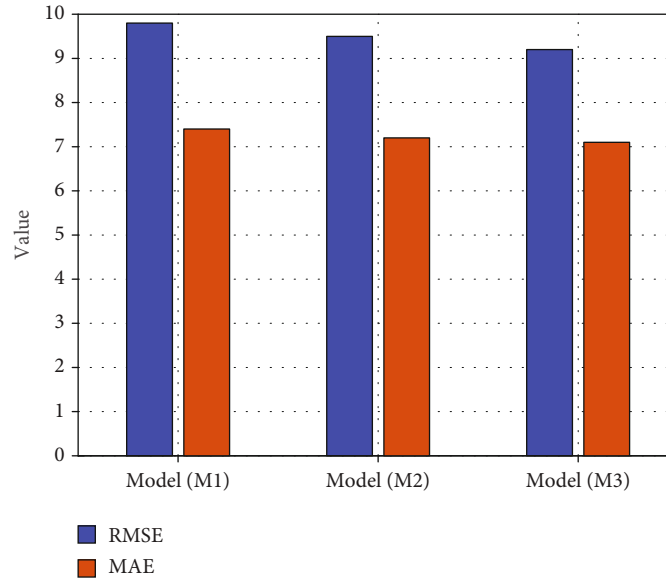


FIGURE 4: The recognition results of the three models in the network.

The depression recognition system designed in this paper is mainly divided into three modules: user voice recording module, depression recognition module, and About Us. The functional module is shown in Figure 6. Different modules corresponding to different functional requirements are as follows:

- (1) User voice recording module: the main function of this module is to realize the function of recording and storing the user's own voice files. The corresponding functional operations can be realized by clicking different button controls. This module is mainly to realize the collection of user voice signal materials for subsequent depression recognition
- (2) Depression recognition module: after the user selects the voice file to be recognized, the selected voice file is recognized for depression based on the trained algorithm model and the result is returned to the interface. The module also realizes the function of playing the selected voice file, so that the user can clearly know the sound content of the selected voice file

TABLE 2: Comparison of experimental results.

Model	Feature	MAE	RMSE
LR	MFCCs	8.112	10.314
	V1	8.240	10.453
CNN	MFCCs	7.984	10.168
	V1	8.067	10.282
DCNN	MFCCs	7.922	9.978
	V1	7.913	9.966
AudioNet	MFCCs	7.785	9.801
	V1	7.504	9.717
Network used	V1+V2+V3	7.230	9.215

- (3) About our module: this part is mainly for product introduction and function description of the speech-based depression recognition application system. It is the software manual to help users understand the main functions and practical uses of the system

TABLE 3: Recognition results of noisy datasets by models.

Noise\model	Index	LR	CNN	DCNN	AudioNet	Network used
Mean 0, variance 0.01	MAE	8.382	8.764	8.226	8.251	7.976
	RMSE	11.103	11.180	10.375	10.469	9.893
Mean 0, variance 0.02	MAE	8.921	9.082	8.886	9.335	8.007
	RMSE	11.872	12.127	10.932	10.903	10.432
Mean 0, variance 0.03	MAE	9.644	9.828	9.273	10.272	8.865
	RMSE	12.824	12.794	11.412	11.583	11.224
Mean 0, variance 0.04	MAE	10.720	10.532	10.023	11.091	9.942
	RMSE	13.983	13.627	12.349	12.622	12.180
Mean 0, variance 0.05	MAE	11.365	11.284	10.892	12.022	10.957
	RMSE	14.921	14.532	13.200	14.177	13.284

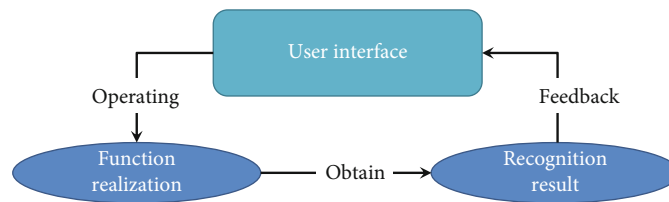


FIGURE 5: Analysis of system design requirements.

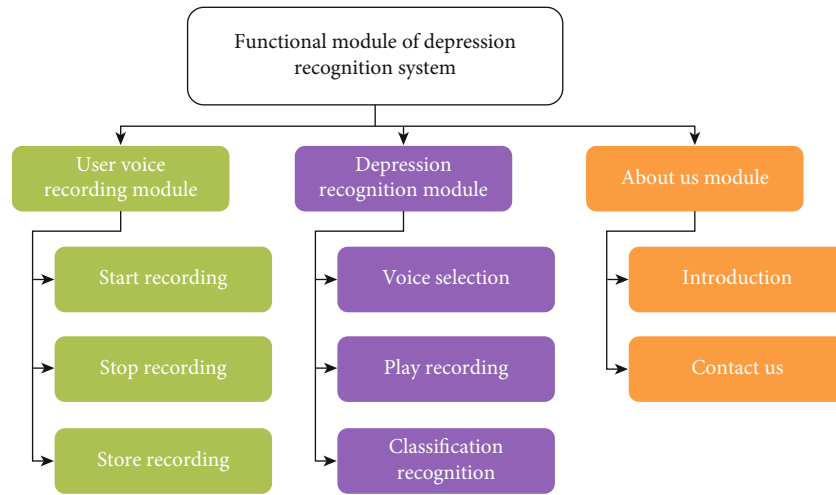


FIGURE 6: System function module.

## 6. Conclusion

This paper designs an intelligent depression recognition system. First, an improved network is used to recognize depression in voice data. The network uses the MADN feature extraction method to reduce the impact of the personalized voice characteristics carried by the sample on the sample. And use the adjacent voice features of the current voice segment to optimize the upper-level model to improve the recognition ability and generalization of the entire network. Experiments have verified that the network used can achieve a better recognition effect, and it is feasible and effective. Secondly, based on this basic research, this paper designs a depression recognition human-computer interaction system. After inputting the user's voice data, the result of depression

recognition is output. The visual interface improves the patient's satisfaction with the visit and can effectively assist the doctor in the diagnosis. In the near future, we will continue to study the role of other modal data in the recognition of depression to further improve the accuracy of identifying depression. And we will also consider to use multiview learning, transfer learning, and the other advanced machine learning technologies to further develop the performance of the proposed intelligent recognition of depression system.

## Data Availability

The labeled dataset used to support the findings of this study is available from the corresponding author upon request.

## Conflicts of Interest

The authors declare no conflicts of interest.

## Acknowledgments

This work was supported by the Jiangsu Province University Philosophy and Social Science Research 2019 Major Project, “Human-computer interaction design research based on artificial intelligence technology” (Project No. 2019 SJZDA118); the Higher Education Research Project of Nanjing Institute of Engineering in 2020, “Research on Cultivating Path of Artificial Intelligence Design Applied Talents” (Project No. 2020YB17); Youth Fund for Humanities and Social Science Research Project of the Ministry of Education (Project No. 20YJC760030); and the National Key R&D Program of China (Grant No. 2017YFB0202303).

## References

- [1] J.-P. Lépine and M. Briley, “The increasing burden of depression,” *Neuropsychiatric Disease and Treatment*, vol. 7, Supplement 1, pp. 3–7, 2011.
- [2] V. Manicavasagar, *A review of depression diagnosis and management*, vol. 34, no. 1, 2012. In *Psych: The Bulletin of the Australian Psychological Society Ltd*, 2012.
- [3] J. Olesen, A. Gustavsson, M. Svensson et al., “The economic cost of brain disorders in Europe,” *European Journal of Neurology*, vol. 19, no. 1, pp. 155–162, 2012.
- [4] W. F. Stewart, J. A. Ricci, E. Chee, S. R. Hahn, and D. Morganstein, “Cost of lost productive work time among US workers with depression,” *Journal of the American Medical Association*, vol. 289, no. 23, pp. 3135–3144, 2003.
- [5] A. J. Mitchell, A. Vaze, and S. Rao, “Clinical diagnosis of depression in primary care: a meta-analysis,” *Lancet*, vol. 374, no. 9690, pp. 609–619, 2009.
- [6] I. Schumann, A. Schneider, C. Kantert, B. Lowe, and K. Linde, “Physicians’ attitudes, diagnostic process and barriers regarding depression diagnosis in primary care: a systematic review of qualitative studies,” *Family Practice*, vol. 29, no. 3, pp. 255–263, 2012.
- [7] T. Inoue, Y. Inagaki, T. Kimura, and O. Shirakawa, “Prevalence and predictors of bipolar disorders in patients with a major depressive episode: the Japanese epidemiological trial with latest measure of bipolar disorder (JET-LMBP),” *Journal of Affective Disorders*, vol. 174, pp. 535–541, 2015.
- [8] T. Vos, C. Allen, M. Arora et al., “Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015,” *Lancet*, vol. 388, no. 10053, pp. 1545–1602, 2016.
- [9] A. Nilsson, “Acoustic analysis of speech variables during depression and after improvement,” *Acta Psychiatrica Scandinavica*, vol. 76, no. 3, pp. 235–245, 1987.
- [10] A. J. Flint, S. E. Black, I. Campbell-Taylor, G. F. Gailey, and C. Levinton, “Abnormal speech articulation, psychomotor retardation, and subcortical dysfunction in major depression,” *Journal of Psychiatric Research*, vol. 27, no. 3, pp. 309–319, 1993.
- [11] J. C. Mundt, P. J. Snyder, M. S. Cannizzaro, K. Chappie, and D. S. Gerlats, “Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology,” *Journal of Neurolinguistics*, vol. 20, no. 1, pp. 50–64, 2007.
- [12] J. C. Mundt, A. P. Vogel, D. E. Feltner, and W. R. Lenderking, “Vocal acoustic biomarkers of depression severity and treatment response,” *Biological Psychiatry*, vol. 72, no. 7, pp. 580–587, 2012.
- [13] M. Cannizzaro, B. Harel, N. Reilly, P. Chappell, and P. J. Snyder, “Voice acoustical measurement of the severity of major depression,” *Brain and Cognition*, vol. 56, no. 1, pp. 30–35, 2004.
- [14] D. J. France, R. G. Shiavi, S. Silverman, M. Wilkes, and M. Silverman, “Acoustical properties of speech as indicators of depression and suicidal risk,” *IEEE transactions on Biomedical Engineering*, vol. 47, no. 7, pp. 829–837, 2000.
- [15] S. Scherer, G. Stratou, G. Lucas et al., “Automatic audiovisual behavior descriptors for psychological disorder analysis,” *Image and Vision Computing*, vol. 32, no. 10, pp. 648–658, 2014.
- [16] S. Kuny and H. H. Stassen, “Speaking behavior and voice sound characteristics in depressive patients during recovery,” *Journal of psychiatric research*, vol. 27, no. 3, pp. 289–307, 1993.
- [17] F. Tolkmitt, H. Helfrich, R. Standke, and K. R. Scherer, “Vocal indicators of psychiatric treatment effects in depressives and schizophrenics,” *Journal of Communication Disorders*, vol. 15, no. 3, pp. 209–222, 1982.
- [18] H. Jiang, B. Hu, Z. Liu et al., “Investigation of different speech types and emotions for detecting depression using different classifiers,” *Speech Communication*, vol. 90, pp. 39–46, 2017.
- [19] S. Gao, V. D. Calhoun, and J. Sui, “Machine learning in major depression: from classification to treatment outcome prediction,” *CNS neuroscience & therapeutics*, vol. 24, no. 11, pp. 1037–1052, 2018.
- [20] J. Lorenzo-Trueba, G. Eje Henter, S. Takaki, J. Yamagishi, Y. Morino, and Y. Ochiai, “Investigating different representations for modeling and controlling multiple emotions in DNN-based speech synthesis,” *Speech Communication*, vol. 99, pp. 135–143, 2018.
- [21] J. F. Zhao, X. Mao, and L. J. Chen, “Learning deep features to recognise speech emotion using merged deep CNN,” *IET Signal Processing*, vol. 12, no. 6, pp. 713–721, 2018.
- [22] A. Jan, H. Meng, Y. F. B. A. Gaus, and F. Zhang, “Artificial intelligent system for automatic depression level analysis through visual and vocal expressions,” *IEEE transactions on autonomous mental development*, vol. 10, no. 3, pp. 668–680, 2018.
- [23] W. Pan, J. Wang, T. Liu et al., “Depression recognition based on speech analysis,” *Kexue Tongbao/Chinese Science Bulletin*, vol. 63, no. 20, pp. 2081–2092, 2018.
- [24] Z. Wang, L. Chen, L. Wang, and G. Diao, “Recognition of audio depression based on convolutional neural network and generative antagonism network model,” *IEEE Access*, vol. 8, pp. 101181–101191, 2020.
- [25] L. He and C. Cao, “Automated depression analysis using convolutional neural networks from speech,” *Journal of Biomedical Informatics*, vol. 83, pp. 103–111, 2018.

## Research Article

# A Violation Information Recognition Method of Live-Broadcasting Platform Based on Machine Learning Technology

Xiaoying Shen <sup>1</sup> and Chao Yuan <sup>2</sup>

<sup>1</sup>Wuxi Professional College of Science and Technology, No. 8 Xixi Road, Wuxi, Jiangsu 214000, China

<sup>2</sup>School of Design, Jiangnan University, 1800 Lihu Avenue, Wuxi, Jiangsu 214122, China

Correspondence should be addressed to Chao Yuan; [circle@jiangnan.edu.cn](mailto:circle@jiangnan.edu.cn)

Received 23 April 2021; Revised 13 May 2021; Accepted 17 May 2021; Published 26 May 2021

Academic Editor: Shan Zhong

Copyright © 2021 Xiaoying Shen and Chao Yuan. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the development of the live broadcast industry, security issues in the live broadcast process have become increasingly apparent. At present, the supervision of various live broadcast platforms is basically in a state of human supervision. Manpower supervision is mainly through user reporting and platform supervision measures. However, there are a large number of live broadcast rooms at the same time, and only relying on human supervision can no longer meet the monitoring needs of live broadcasts. Based on this situation, this study proposes a violation information recognition method of a live-broadcasting platform based on machine learning technology. By analyzing the similarities and differences between normal live broadcasts and violation live broadcasts, combined with the characteristics of violation image data, this study mainly detects human skin color and sensitive parts. A prominent feature of violation images is that they contain a large area of naked skin, and the ratio of the area of naked skin to the overall image area of the violation image will exceed the threshold. Skin color recognition plays a role in initial target positioning. The accuracy of skin color recognition is directly related to the recognition accuracy of the entire system, so skin color recognition is the most important part of violation information recognition. Although there are many effective skin color recognition technologies, the accuracy and stability of skin color recognition still need to be improved due to the influence of various external factors, such as light intensity, light source color, and physical equipment. When it is detected that the area of the skin color in the live screen exceeds the threshold, it is preliminarily determined to be a suspected violation video. In order to improve the recognition accuracy, it is necessary to detect sensitive parts of the suspected video. Naked female breasts are a very obvious feature in violation images. This study uses a chest feature extraction method to detect the chest in the image. When the recognition result is a violation image, it is determined that the live broadcast involves violation content. The machine learning algorithm is simple to implement, and the parameters are easy to adjust. The classifier training requires a short time and is suitable for live violation information recognition scenarios. The experimental results on the adopted data set show that the method used in this article can effectively detect videos with violation content. The recognition rate is as high as 85.98%, which is suitable for a real-life environment and has good practical significance.

## 1. Introduction

In recent years, the webcast industry has developed rapidly, and it has become the main channel of entertainment for the majority of netizens. The blowout-style webcast also brought chaos in the live broadcast. Some webcast platforms lack a sense of responsibility and do not have full-time platform content supervisors. The anchors of the live broadcast platform also lack self-discipline and have a shallow legal awareness, which has caused an endless stream of undesir-

able phenomena such as online violence and online obscenity. During the live broadcast, there are phenomena such as pornography, gambling, drugs, and sneak shots, which seriously endanger the physical and mental health of netizens. At the same time, this has also caused a serious negative impact on the good development of society. Due to the large number of live broadcast platforms, the wide coverage of content, and the inconsistent live broadcast time, it is obviously incapable of relying solely on network police to monitor. Therefore, intelligent violation information recognition

technology needs to be introduced into the field of webcasting. Currently, violation information recognition is mainly carried out from the following aspects. One is skin color recognition. The core principle is to determine that the image contains violation content after detecting that the area of the skin in the screen exceeds a prescribed threshold. When multiple consecutive images have the same situation, it is determined that the video is a violation video. Reference [1] used skin color segmentation and geometric features of human pose to identify sensitive images. The basic idea is to treat the human body as a combination of several columnar regions according to certain rules. However, its disadvantages are that it can only process a single type of image, lack of adaptability and versatility, slow processing speed, and low image recognition rate. Reference [2] proposed the use of skin color recognition, texture analysis, and feature vector classification to identify sensitive pictures. Since a simple skin color model is used, the judgment result is largely dependent on the result of skin region extraction, which has great limitations. Reference [3] uses key frames to build an XYZ model. The recognition of a single frame is taken as the X axis, the framing is the Y axis, and the X and Y axes are parameters to construct the Z axis to determine whether the frame contains sensitive images. Many subsequent studies have focused on improving the recognition rate of human skin color [4–6]. The second is human behavior recognition. Human behavior recognition is to process time-varying data, that is, to select the parameters used to describe the static posture of the human body from a series of key frames in a time sequence obtained from the image sequence, connect them to a set of parameters, and match them with the predefined action template, so that the computer can describe the behavior of the human body in natural language. There are two common human behavior recognition technologies in videos: (1) A method based on model matching. Yuliie et al. proposed a method based on flexible templates to extract facial features. Sakai et al. used sub-models of facial features and face contours to detect frontal faces in images. Craw et al. proposed a recognition method based on frontal faces. First, extract the edges, then connect the edges, detect the face template library according to the relevant conditions, and finally use the same method and different sizes to search for the facial features. (2) A method based on state space. Reference [7] proposed a method of describing behaviors with optical flow direction histograms, which opened up a new way for human behavior recognition. The third is based on the recognition of sensitive parts. Sensitive parts mainly refer to parts such as breasts. This method is mainly reflected in the training of the recognition model. Commonly used recognition models mainly include support vector machine (SVM) [8–9], AdaBoost model [10–12], and neural network [13–15]. The fourth is the recognition technology based on the human face. The main part of violation images is undoubtedly the human body, and the human face is also an important part of the human body. In violation images, except for individual violation images which are closed-up images of sensitive parts, most

of them contain facial information. Therefore, accurate face recognition can also be used as one of the judging features of violation images. Face recognition is mainly divided into three categories: (1) Methods based on skin color [16–17]. Similar to skin color recognition, the judgment is based on the difference between the color of the face skin and the background. This method requires relatively high color. (2) Feature-based methods [18–20]. It is also judged based on all the texture features of the face. This method is less sensitive to color, and at the same time, it is more robust to the recognition of various skin colors. (3) Methods based on statistics [21–22]. Count the face image data and then design a classifier to classify and judge the images to be detected.

It can be seen from the above research that most types of violation recognition need to design a classifier with good performance. Currently, classifiers mainly include two types based on machine learning and deep learning. In the field of machine learning, the commonly used classification models mainly include the Gaussian model [23–24], SVM [25–26], AdaBoost [27–28], and fuzzy system [29–30]. The use of machine learning algorithms often needs to be equipped with suitable feature extraction methods to get the desired results. Commonly used feature extraction methods mainly include wavelet decomposition and principal component analysis [31–32]. In the field of deep learning, the commonly used classification models mainly include convolutional neural networks [33–34], cyclic neural networks [35–36], and long- and short-term memory neural networks [37]. This article proposes a violation information recognition method based on machine learning for the scene of a live webcast. The method can realize real-time rapid recognition and is suitable for the field of webcasting. The main work of this study is summarized as follows:

- (1) For the recognition of violation information in live broadcasts, skin color recognition is first used for preliminary filtering. Specifically, because the YCbCr color space is simple and it is easy to separate the brightness components in the image, this study uses the Gaussian model to train in the YCbCr color space to obtain a skin color recognition model. According to the trained Bayesian YCbCr skin color model, the test images to be tested are tested. The test image is converted from RGB space to YCbCr space, and its YCbCr component is taken. This component is input into the Bayesian classification model, and the skin color segmentation result can be obtained
- (2) Due to the error in the skin color recognition result, in order to improve the recognition accuracy, a second round of recognition is performed on the suspected video data. Specifically, Haar-like features are extracted as female breast recognition features. Haar-like features are used to describe the gray features of local areas. In the training window of  $40 * 40$  pixels, train and select

a weak classifier with strong classification ability. Each weak classifier is combined into a strong classifier according to a certain method, and then, a cascade classifier is obtained by training. Finally, the cascade classifier obtained by training is used to detect the test samples with a  $40 * 40$  pixel sliding window

- (3) In order to verify the recognition effect of the method in this study on the violation content of the live broadcast platform, the data set is used for experimental comparison. In order to illustrate the effectiveness of the two stages of recognition in this article, the experiment compares the recognition of skin color only, the recognition of sensitive parts, and the cascade recognition of the two. In order to verify the applicability and superiority of the classification model selected at each stage, the experimental part also made a comparative analysis of the classification model. The experimental results show that the method in this study has a good recognition effect and practical value

## 2. Related Work

*2.1. Violation Information Recognition Based on Skin Color.* Violation image recognition belongs to image recognition. Due to the complex image background and the susceptibility to illumination, as well as the diversity of the target human body's gestures, it is difficult to use a single model to represent the characteristics of violation images. A prominent feature of violation images is the large area of bare skin. Various studies have shown that the basis of violation image recognition technology is the extraction of the skin area in the image. Firstly, it is judged whether the pixels in the image are skin pixels; on this basis, the next step of judgment and recognition is carried out, and finally, the whole image is judged.

Skin color recognition technology is mainly used in face recognition, human body recognition, and violation image identification. For violation images, the most important thing is to detect the naked human skin in the image. Extensive nudity is the most important feature of violation images. The ratio of the naked skin area of the violation image to the overall image area must exceed a certain threshold. The current mainstream violation image recognition technologies are based on skin color recognition, and subsequent analysis and judgment are made on this basis. Skin color recognition has played a role in locating the target initially. The accuracy of skin color recognition is directly related to the recognition accuracy of the entire system, so skin color recognition is the most important part of the entire violation content recognition system. Many researchers have discussed the skin color recognition technology. Due to the influence of various external factors, such as light intensity, light source color, and physical equipment, it is difficult to determine skin color simply and accurately. So far, there is no stable and effective method to determine the part of the skin in the image. Most of the current research is devoted to improving the accuracy and stability of the algorithm. The violation video recognition process based on skin color is shown in Figure 1:

As shown in Figure 1, the collected live video is first decomposed into one frame by frame image. Second, the skin color feature of the image is extracted. Third, the skin area is calculated. When the area is greater than the set threshold, the image is determined to be an abnormal image. When the area is less than the threshold, the image is determined to be a normal image. Finally, when multiple frames of continuous images are judged to be abnormal images, the video is considered to contain violation content.

*2.2. Violation Information Recognition Based on Human Face.* Studies have shown that the proportion of human faces in normal images to skin color is much greater than that in violation images. Therefore, after adding face recognition to judge the proportion of skin color occupied by human faces, it can reduce the false recognition rate of normal images and basically does not affect the judgment of violation images. There are many researches on face recognition, but each research has its shortcomings. At present, the most practically used face recognition method is based on the AdaBoost algorithm. This article uses Viola's face recognition method, which is based on a cascade detector. The face recognition process based on cascaded detectors is shown in Figure 2.

As shown in Figure 2, the Haar-like feature is first used to characterize the face. In order to accelerate the calculation of the feature value, the "integral graph" technology is applied in the calculation process. Second, a series of weak classifiers that are most suitable for face determination are selected through the AdaBoost algorithm. Combine multiple weak classifiers into one strong classifier based on voting. Finally, multiple strong classifiers are connected in series to form an efficient stacked classifier with a cascade structure.

The general operation steps for specific implementation are as follows:

- (1) Extract features from a  $20 * 20$  picture
- (2) Thousands of cut face pictures and tens of thousands of background pictures are used as training samples. Usually, the training images are uniformly scaled to a  $20 * 20$  square. At this size, the number of Haar features that can be used is about 10,000. Then, use the AdaBoost algorithm to select thousands of effective features from these tens of thousands of features to construct a classifier
- (3) After the face detector has been trained, it can be used. Before recognition, the image is scaled according to a certain ratio. On the zoomed picture, a  $20 * 20$  subwindow is used to sequentially determine whether it is a human face or a nonhuman face

*2.3. Violation Information Recognition Based on Sensitive Parts.* Nudity is an important visual feature of violation images, and the recognition of violation shots in sensitive parts is more specific. According to the texture characteristics of sensitive parts, a classifier of sensitive parts is trained to detect pornography, which can detect violation content more accurately. The training process of the sensitive part classifier refers to the process of training the classifier of the AdaBoost

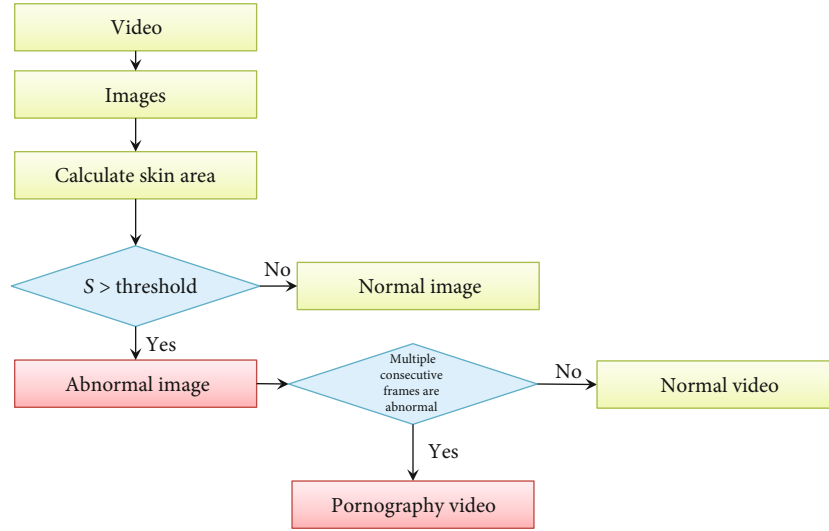


FIGURE 1: Violation video recognition process based on skin color.

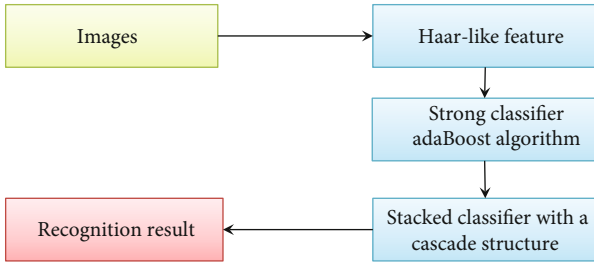


FIGURE 2: Face recognition process based on cascade detector.

algorithm in face recognition. The core steps of the sensitive part recognition based on AdaBoost learning are as follows: First, the binary image extracted by skin color recognition, face recognition, and connected domains is segmented to obtain the region to be inspected. Then, the sensitive part classifier recognition of the sliding window is performed, and if it fails, it is judged as a normal image. Then, perform the areola classifier recognition of the sliding window; if it fails, it is still judged as a normal video. If it passes, it is judged as a suspected violation image. The process is shown in Figure 3, where breast classifier recognition and areola classifier recognition are the core of the method.

### 3. Violation Information Recognition on Live Video

**3.1. Recognition Framework.** Figure 4 shows the violation information recognition framework used in this article. First, analyze and preprocess the input image. This work is mainly to pave the way for subsequent image analysis and individual feature recognition. The preprocessed image is subjected to the skin color recognition in the first step, and the skin color recognition result is obtained. When the result is a suspected violation image, the image is input to the second stage of sensitive part recognition. When the recognition result of the sensitive part is a violation image, the image is determined

to be a violation image. When multiple adjacent frames of images are determined to be violation images, it is determined that the live video contains violation content.

**3.2. Bayesian Skin Color Recognition.** The skin color recognition process based on Bayesian is shown in Figure 5. First, the skin color part and the nonskin color part are extracted manually from each sample image. Second, the probability map of the skin color and nonskin color parts of the sample image is calculated. After manual segmentation, the histogram statistics are performed on the CbCr component and normalized. In the statistical process of the probability map, the RGB value of each pixel in the skin color part and the nonskin color part is first converted to the YCbCr space. Then, accumulate the skin color and nonskin color of the CbCr components and normalize them. Two-dimensional probability statistics of skin color and nonskin color of the CbCr component are obtained. Finally, perform Bayesian classification calculation on  $225 * 225$  CbCr and obtain the Bayesian classification model of skin color on the CbCr component.

This study chooses YCbCr as the color space for skin color recognition. The first reason is that YCbCr is widely used in many vision technologies. YCbCr has a composition principle similar to the human visual perception process and can be directly applied to image formats such as JPEG or MPEG without having to convert it into other color spaces. The second reason is that the YCbCr color space has the advantage of separating the brightness components in the image. The distribution of skin pixels on the YCbCr channel is relatively concentrated and has good cohesion. The third reason is that the YCbCr color space coordinate representation is simpler than other color spaces.

Bayesian method [38] is a method used to solve statistical problems based on Bayesian theory. It can well connect the prior probability and posterior probability of an event. For skin color recognition, the unknown sample  $X$  is first classified into a skin color category  $w_1$  or a nonskin color category  $w_2$ . Let  $C_{ij}$  denote the cost of  $X$  at  $w_1$  being designated as  $w_1$ .

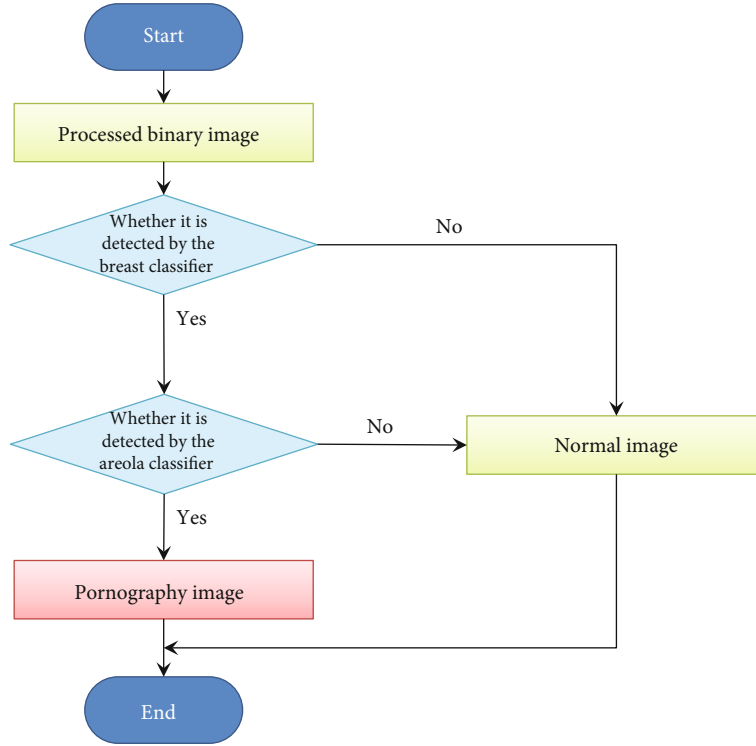


FIGURE 3: Sensitive part recognition process.

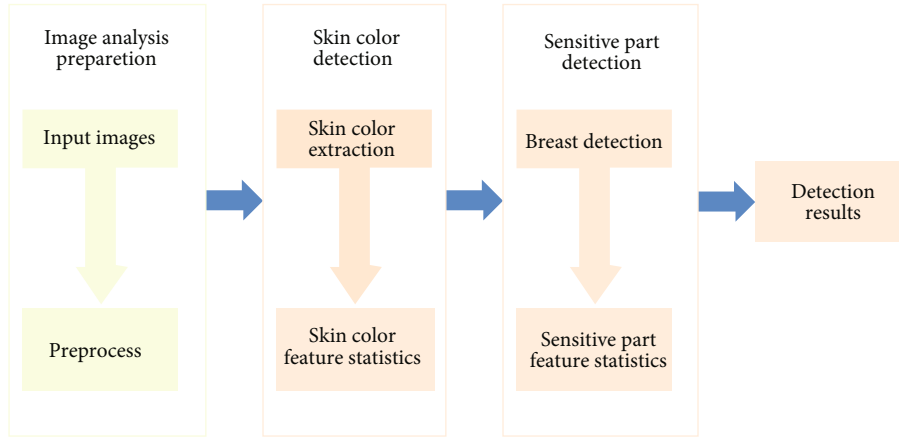


FIGURE 4: Recognition framework diagram.

When  $i = j$ , the classification is correct, that is, it is correctly detected whether the  $X$  is a skin pixel. At this time,  $C_{ij}$  is the cost of correct classification. When  $i \neq j$ , it means that the classification is wrong, that is, the skin color point is detected as the nonskin color part, or the nonskin color part is falsely detected as the skin color. At this time,  $C_{ij}$  is the cost of classification when the error is detected. Let the variable  $R_i(X)$  be the total cost borne by the sample  $X$  to be determined as the  $i$  type. According to the above assumptions, we can get

$$R_1 = C_{11} \cdot p(w_1 | X) + C_{12} \cdot p(w_2 | X), \quad (1)$$

$$R_2 = C_{21} \cdot p(w_1 | X) + C_{22} \cdot p(w_2 | X). \quad (2)$$

The  $p(w_1 | X)$  represents the conditional probability that the unknown sample  $X$  belongs to  $w_1$ , which stipulates

$$R_1(X) < R_2(X) \Rightarrow X \in w_1, \quad (3)$$

$$R_1(X) > R_2(X) \Rightarrow X \in w_2. \quad (4)$$

It can be obtained by the above formula:

$$(C_{11} - C_{21}) \cdot p(w_1 | X) < (C_{22} - C_{12}) \cdot p(w_2 | X) \Rightarrow X \in w_1, \quad (5)$$



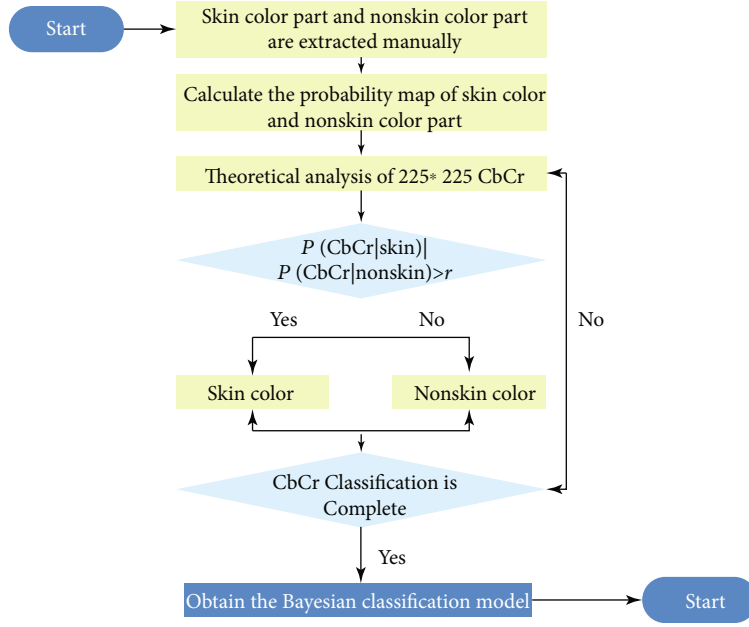


FIGURE 5: Skin color recognition process based on Bayesian.

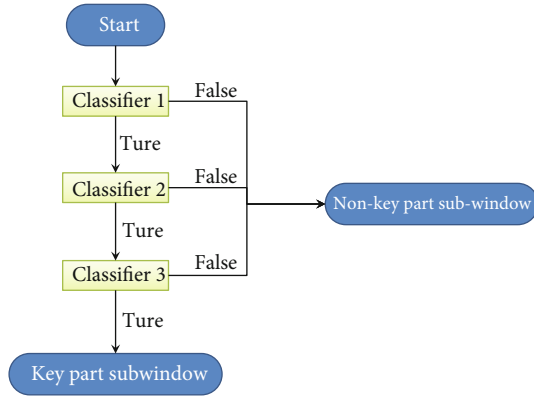


FIGURE 6: AdaBoost classification cascade structure diagram.

$$(C_{11} - C_{21}) \cdot p(w_1 | X) > (C_{22} - C_{12}) \cdot p(w_2 | X) \Rightarrow X \in w_2. \quad (6)$$

Finally, import the Bayesian formula to get

$$p(w_1 | X) = \frac{p(X | w_1)p(w_1)}{p(X)}. \quad (7)$$

After sorting, the following formula is obtained:

$$\frac{p(X | w_1)}{p(X | w_2)} > \tau \Rightarrow X \in w_1, \quad (8)$$

$$\frac{p(X | w_1)}{p(X | w_2)} < \tau \Rightarrow X \in w_2, \quad (9)$$

where

$$\tau = \frac{(C_{22} - C_{12}) \cdot p(w_2)}{(C_{11} - C_{21}) \cdot p(w_1)}. \quad (10)$$

The key of this method is to determine an optimal threshold  $\tau$  to achieve accurate skin color recognition. When  $p(\text{CbCr} | \text{skin})/p(\text{CbCr} | \text{nonskin}) > \tau$ , this type of CbCr is a skin color point, and vice versa, it is a nonskin color point. According to reference [39], the optimal value of  $t$  is between 2 and 4. This article uses a single-point skin color recognition method. Extract the color information of CbCr from a single point, and then, make a decision based on the statistical histogram information. This method achieves better results and is faster at the same time.

**3.3. Sensitive Part Recognition Based on AdaBoost.** The Haar-like feature is extracted as the main feature for female breast recognition. The Haar-like feature is used to describe the gray features of local areas. This breast feature extraction method easily misjudges the eyes, belly button, and other parts as breasts. In order to avoid this problem as much as possible, the red color of the chest color is considered stronger. Let the color characteristic value of the chest be  $N^0 = N_r - N_b$ .  $N_r$  is the information difference between the R channel and the B and G channels.  $N_b$  is the information difference between the B and G channels. The calculation formulas of  $N_r$  and  $N_b$  are as follows:

$$N_r = \begin{cases} V = 2r - g - b (V > 0), \\ 0 (V \leq 0), \end{cases} \quad (11)$$

TABLE 1: Introduction to the data set used.

Video category	Number of videos	Duration	Average number of frames extracted in each video
Pornography	400	57	15.6
Nonviolation (simple)	200	11.5	33.8
Nonviolation (difficult)	200	8.5	17.5
All	800	77	20.6

$$N_b = \begin{cases} V = b - g(V > 0), \\ 0(V \leq 0), \end{cases} \quad (12)$$

$$N^0 = N_r - N_b, \quad (13)$$

where  $r$ ,  $g$ , and  $b$ , respectively, represent the component information value of the pixel on the Red, Green, and Black (RGB) channel. The calculation formula for the average eigenvalue of the target area is as follows:

$$\bar{N}^0 = \frac{1}{n} \sum N^0. \quad (14)$$

The  $n$  is the number of pixels in the target area. When  $\bar{N}^0$  is greater than the set threshold, it can be considered that the red color feature here is large, and misrecognition areas such as the eyes and belly buttons can be excluded. At the same time, the calculated area is the target area for classification, and the area is small, so the calculation process will not affect the speed.

In a training window of  $40 \times 40$  pixels, a weak classifier with strong classification ability is trained. According to a certain method, a strong classifier is combined into a strong classifier, and then, a cascaded classifier is trained. Finally, the trained cascade classifier and a  $40 \times 40$  pixel sliding window are used to detect the test samples. Figure 6 shows the structure of the AdaBoost cascade classifier.

## 4. Experiment and Analysis

*4.1. Experimental Data.* The experimental data set adopted in this article is a public video data set. The specific representation of this data set is shown in Table 1. Figure 7 is an example of this adopted data set.

*4.2. Experimental Environment.* To analyze the performance of each method, the evaluation indicators used in this study are precision, recall, and F1-measure (F1). The calculation formula of each evaluation index is as follows:

$$\text{precision} = \frac{TP}{TP + FP}, \quad (15)$$

$$\text{recall} = \frac{TP}{TP + FN}, \quad (16)$$

$$F1 = \frac{(w \times w + 1) \times \text{precision} \times \text{recall}}{(w \times w) \times \text{precision} + \text{recall}}, \quad (17)$$

where TP stands for the number of positive classes pre-

dicted as positive classes, TN stands for negative class predictions as negative classes, FP stands for negative class predictions as positive classes, and FN stands for positive class predictions as negative classes. The  $w$  in Equation (17) is the weight value. When the value of  $w$  is greater than 1, the accuracy rate is more important, and when the value of  $w$  is less than 1, the recall rate is more important. In this study, the value of  $w$  is set to 1, indicating that the requirements for precision and recall are the same. The values of the above three indicators are all between 0 and 1. The closer the value is to 1, the better the classification effect of the model.

The hardware configuration information used in this experiment is as follows: CPU is Intel Core i7, graphics memory is GTX960M 4G, and memory is 16G. The operating system is Windows10 64-bit, and the development language is MATLAB.

*4.3. Experimental Results and Analysis.* The experiment mainly conducts comparative analysis from three aspects. One is to perform skin color recognition only, the other is to detect sensitive parts only, and the third is to combine the two recognition methods.

*4.3.1. Skin Color Recognition Experiment.* To compare the effectiveness of the Bayesian model for skin color recognition in this study, the comparison model uses artificial neural networks (ANN) [40] and SVM [41]. The experimental results are shown in Table 2:

It can be seen from the experimental results that the skin color recognition gap based on the three models is not large. The Bayesian classification model based on YCbCr space has achieved the best recognition effect. Because the distribution of skin pixels on the CbCr component has good cohesion, the Bayesian classification model based on YCbCr space can achieve better skin color recognition results. This method is simple and fast. The classification result based on the Bayesian criterion is also more accurate. The experimental results in this study are in line with expectations. However, this method strongly depends on the training sample images, and the comprehensive and reasonable selection of the sample library has a greater impact on the results. At the same time, this method is a static segmentation method. After the classifier is determined, the segmentation threshold of the test image cannot be flexibly changed. Therefore, if the judgment condition is adjusted, when the judgment condition for skin color is strengthened, the recognition result produced by the trained classifier will make it difficult to detect part of the skin color. And when the skin color judgment

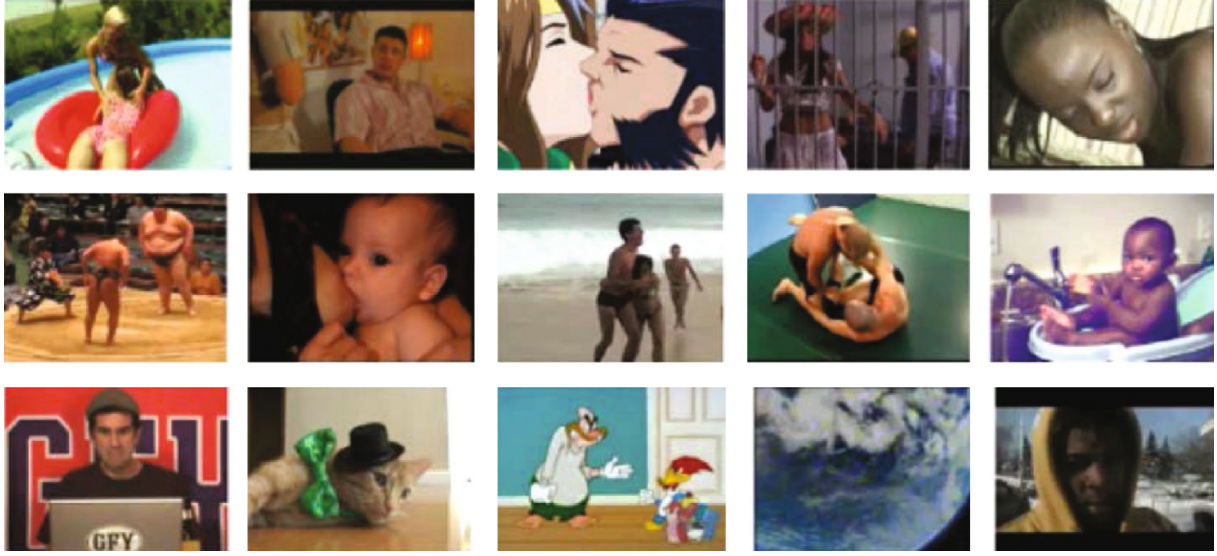


FIGURE 7: Data set legend.

TABLE 2: Skin color recognition results of each model.

Model	Precision (%)	Recall (%)	F1 (%)
ANN	81.21	81.32	81.15
SVM	81.84	81.97	81.48
Bayes	82.85	82.89	82.24

TABLE 3: Sensitive part recognition results of each model.

Model	Precision (%)	Recall (%)	F1 (%)
SVM	78.04	78.92	78.98
TSK	80.65	80.36	80.46
AdaBoost	81.30	81.74	80.62

TABLE 4: Recognition results based on the proposed framework.

Recognition method	Precision (%)	Recall (%)	F1 (%)
Skin color recognition	82.85	82.89	82.24
Sensitive part recognition	81.30	81.74	80.62
Proposed framework	85.98	86.20	85.75

condition is relaxed, some background points will be mistakenly detected as skin color points.

**4.3.2. Sensitive Part Recognition Experiment.** Sensitive part recognition is performed on the original data. The comparison models include the SVM and TSK fuzzy systems [42]. The results of sensitive part recognition by each model are shown in Table 3.

The data in Table 3 shows that the recognition effect of sensitive parts based on the AdaBoost model is obviously better. The Haar feature classifier has a high positive recognition rate and a high false recognition rate. Many nontarget areas are also mistaken for sensitive parts. Appropriately increasing the number of classifier stages can reduce the false recog-

niton rate, but the false recognition rate is still high. At the same time, an increase in the number of classifier stages will also lead to an increase in computing time. Compared with face recognition, the characteristics of the areola are not obvious, except that the color of the center of the area is darker than the surrounding area, and there is no more obvious feature available. This is why the overall accuracy of recognition of sensitive parts is lower than that of skin color recognition.

**4.3.3. Recognition Experiment Based on the Proposed Framework.** To improve the recognition accuracy, this research proposes to use multistep recognition. First, the original sample is input to the skin color recognition module, and the suspected violation sample is detected. Second, input the suspected sample to the sensitive part recognition module. The test results of the framework proposed in this study are shown in Table 4. It can be seen from the experimental results that the framework proposed in this article greatly improves the recognition accuracy of violation content. This is because the recognition of multiple links can reduce the false recognition rate.

## 5. Conclusion

Aiming at the recognition of violation content in live broadcast scenes, this study proposes a recognition framework based on skin color and sensitive parts. In the skin color recognition link, a Bayesian recognition method based on the YCbCr space is used. However, relying only on skin color recognition results as the only criterion for violation content recognition cannot meet the needs of practical applications. In the judgment of images such as wearing sexy celebrity portraits or avatar ID photos, since such pictures also have a large amount of naked skin, the judgment will have a large false recognition rate. Therefore, a sensitive part recognition link is added. This session uses an AdaBoost-based recognition method for bare breasts. The experimental results show that the recognition framework proposed in this study can

effectively detect whether the live broadcast scene contains violation content. This research has very good practical value. To shorten the recognition time, this research omitted the face recognition link. But at the same time, the false recognition rate is increased. The next step is to improve the recognition accuracy without increasing the recognition time as much as possible. This research intends to introduce deep learning to extract deep features of images to improve recognition efficiency.

## Data Availability

The labeled data set used to support the findings of this study is available from the corresponding author upon request.

## Conflicts of Interest

The authors declare no conflicts of interest.

## Acknowledgments

This work was supported in part by the National Social Science Fund Youth Project under Grant 17BGL102, Excellent Project of Jiangsu Province Social Science Union under Grant 15SYC-043, Soft Science Research of Wuxi Science and Technology Association under Grant KX15-B-01, and Fundamental Research Funds for the Central Universities under Grant 2015ZX18.

## References

- [1] D. A. Forsyth and M. M. Fleck, "Identifying nude pictures," in *Proc. of the 3rd IEEE Workshop on Applications of Computer Vision*, pp. 103–108, Sarasota, FL, USA, 1996.
- [2] C. Y. Jeong, J. S. Kim, and K. S. Hong, "Appearance-based nude image recognition," in *Proceedings of the 17th International Conference on Pattern Recognition*, pp. 1051–4651, Cambridge, UK, 2004.
- [3] H. Lee, S. Lee, and T. Nam, "Implementation of high performance objectionable video classification system," in *2006 8th International Conference Advanced Communication Technology*, no. 2, pp. 20–22, Phoenix Park, Korea (South), 2006.
- [4] D. Brown, I. Craw, and J. Lewthwaite, "A SOM based approach to skin recognition with application in real time system," in *Proc. of British Conf. on Machine Vision*, London, UK, 2009.
- [5] J. Basilio, G. A. Torres, G. S. Pérez, L. Medina, and H. Meana, "Explicit image recognition using YCbCr space color model as skin recognition," in *Proc. 2011 American Conf. on Applied Mathematics and the Fifth WSEAS Int. Conf. on Computer Engineering and Applications World Scientific and Engineering Academy and Society (WSEAS)*, pp. 123–128, Puerto Morelos, Mexico, 2011.
- [6] M. J. Jones and J. M. Reh, "Statistical color models with application to skin recognition," *International Journal of Computer Vision*, vol. 46, no. 1, pp. 81–96, 2002.
- [7] X. Li, "HMM based action recognition using oriented histograms of optical flow field," *Electronics Letters*, vol. 43, no. 10, pp. 560–561, 2007.
- [8] C. H. Miller, M. D. Sacchet, and I. H. Gotlib, "Support vector machines and affective science," *Emotion Review*, vol. 12, no. 4, pp. 297–308, 2020.
- [9] Z. Zhang, Z. Xu, J. Tan, and H. Zou, "Multi-class support vector machine based on the minimization of class variance," *Neural Processing Letters*, vol. 53, no. 1, pp. 517–533, 2021.
- [10] Q. Huang, Y. Chen, L. Liu, D. Tao, and X. Li, "On combining biclustering mining and AdaBoost for breast tumor classification," *IEEE Transactions on Knowledge & Data Engineering*, vol. 32, no. 4, pp. 728–738, 2020.
- [11] G. Gosztolya and R. Busa-Fekete, "Calibrating AdaBoost for phoneme classification," *Soft Computing*, vol. 23, no. 1, pp. 115–128, 2019.
- [12] N. Alam, S. Agarwal, M. Gautam, A. Srivasata, and R. Nijhawan, "A novel AdaBoost approach for cloud type classification," *Journal of Advanced Research in Dynamical and Control Systems*, vol. 11, no. 6Special, pp. 372–376, 2019.
- [13] Y. Yang, L. Chen, Y. Xiong, S. Li, and M. Xu, "Global sensitivity analysis based on BP neural network for thermal design parameters," *Journal of Thermophysics and Heat Transfer*, vol. 35, no. 1, pp. 187–199, 2021.
- [14] S. Yang, X. Zhu, L. Zhang, L. Wang, and X. Wang, "Classification and prediction of Tibetan medical syndrome based on the improved BP neural network," *IEEE Access*, vol. 8, pp. 31114–31125, 2020.
- [15] N. Pelchat and M. Craig, "Neural network music genre classification," *Canadian Journal of Electrical and Computer Engineering*, vol. 43, no. 3, pp. 170–173, 2020.
- [16] S. Naji, H. A. Jalab, and S. A. Kareem, "A survey on skin detection in colored images," *Artificial Intelligence Review*, vol. 52, no. 2, pp. 1041–1087, 2019.
- [17] A. Albiol, L. Torres, and E. J. Delp, "An unsupervised color image segmentation algorithm for face recognition applications," in *International Conference on Image Processing*, pp. 681–684, Thessaloniki, Greece, 2001.
- [18] C. Tao, Y. Duan, and X. Hong, "Discrimination of fabric frictional sounds based on Haar features," *Textile Research Journal*, vol. 89, no. 11, pp. 2067–2074, 2019.
- [19] M. Oualla, K. Ounachad, A. Hnini, and A. Sadiq, "The fast integration of a rotated Haar-like feature for face detection," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, no. 3, pp. 4055–4062, 2020.
- [20] E. Owusu, J.-D. Abdulai, and Y. Zhan, "Face recognition based on multilayer feed-forward neural network and Haar features," *Software: Practice and Experience*, vol. 49, no. 1, pp. 120–129, 2019.
- [21] T. Rajeshkumar, U. Samsudeen, S. Sangeetha, and U. Sudha Rani, "Enhanced visual attendance system by face recognition USING K-nearest neighbor algorithm," *Journal of Advanced Research in Dynamical and Control Systems*, vol. 11, no. 6-Special, pp. 141–147, 2019.
- [22] S. Eko and M. Adharul, "Implementation of K-nearest neighbors face recognition on low-power processor," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 13, no. 3, pp. 949–954, 2015.
- [23] F. Ay, G. İnce, M. E. Kamaşak, and K. Y. Ekşi, "Classification of pulsars with Dirichlet process Gaussian mixture model," *Monthly Notices of the Royal Astronomical Society*, vol. 493, no. 1, pp. 713–722, 2020.
- [24] P. Branislav, K. Jernej, and N. Marko, "Gaussian mixture model based classification revisited: application to the bearing

- fault classification,” *Strojniški Vestnik – Journal of Mechanical Engineering*, vol. 66, no. 4, pp. 215–226, 2020.
- [25] I. Prabhakaran, Z. Wu, C. Lee et al., “Gaussian mixture models for probabilistic classification of breast cancer,” *Cancer Research*, vol. 79, no. 13, pp. 3492–3502, 2019.
- [26] M.-H. Roy and D. Larocque, “Prediction intervals with random forests,” *Statistical Methods in Medical Research*, vol. 29, no. 1, pp. 205–229, 2020.
- [27] Y. Yu, L. Wang, H. Huang, and W. Yang, “An improved random forest algorithm,” *Journal of Physics: Conference Series*, vol. 1646, article 012070, 2020.
- [28] P. Xu, Z. Deng, J. Wang, Q. Zhang, K.-S. Choi, and S. Wang, “Transfer representation learning with TSK fuzzy system,” *IEEE Transactions on Fuzzy Systems*, vol. 29, no. 3, pp. 649–663, 2021.
- [29] Z. Namadchian and A. Zare, “Stability analysis of dynamic nonlinear interval type-2 TSK fuzzy control systems based on describing function,” *Soft Computing*, vol. 24, no. 19, pp. 14623–14636, 2020.
- [30] D. Wu, Y. Yuan, J. Huang, and Y. Tan, “Optimize TSK fuzzy systems for regression problems: minibatch gradient descent with regularization, DropRule, and AdaBound (MBGD-RDA),” *IEEE Transactions on Fuzzy Systems*, vol. 28, no. 5, pp. 1003–1015, 2020.
- [31] M. Cotronei, M. Rossini, T. Sauer, and E. Volontè, “Filters for anisotropic wavelet decompositions,” *Journal of Computational & Applied Mathematics*, vol. 349, pp. 316–330, 2019.
- [32] K. Sando and H. Hino, “Modal principal component analysis,” *Neural Computation*, vol. 32, no. 10, pp. 1–35, 2020.
- [33] X. Wu, Y. Shi, S. Fomel, L. Liang, Q. Zhang, and A. Z. Yusifov, “FaultNet3D: predicting fault probabilities, strikes, and dips with a single convolutional neural network,” *IEEE Transactions on Geoscience & Remote Sensing*, vol. 57, no. 11, pp. 9138–9155, 2019.
- [34] Ş. Öztürk and U. Özkaya, “Skin lesion segmentation with improved convolutional neural network,” *Journal of Digital Imaging*, vol. 33, no. 4, pp. 958–970, 2020.
- [35] M. E. Valle and R. A. Lobo, “Hypercomplex-valued recurrent correlation neural networks,” *Neurocomputing*, vol. 432, pp. 111–123, 2021.
- [36] M. Nikolay and S. Michael, “Target propagation in recurrent neural networks,” *Journal of Machine Learning Research*, vol. 21, no. 1-25, pp. 1–33, 2020.
- [37] G. Van Houdt, C. Mosquera, and G. Nápoles, “A review on the long short-term memory model,” *Artificial Intelligence Review*, vol. 53, no. 8, pp. 5929–5955, 2020.
- [38] Z. Geng, Q. Meng, J. Bai et al., “A model-free Bayesian classifier,” *Information Sciences*, vol. 482, pp. 171–188, 2019.
- [39] D. Chai and A. Ouzerdoum, “A Bayesian approach to skin color classification in YCbCr color space,” in *2000 TENCON Proceedings. Intelligent Systems and Technologies for the New Millennium (Cat. No.00CH37119)*, pp. 421–424, Kuala Lumpur, Malaysia, 2000.
- [40] F. Y. Dalkiran and M. Toraman, “Predicting thrust of aircraft using artificial neural networks,” *Aircraft Engineering and Aerospace Technology*, vol. 93, no. 1, pp. 35–41, 2021.
- [41] M. Malarvel and H. Singh, “An autonomous technique for weld defects detection and classification using multi-class support vector machine in X-radiography image,” *Optik*, vol. 231, article 166342, 2021.
- [42] Y. Jiang, Y. Zhangy, C. Lin, D. Wu, and C.-T. Lin, “EEG-based driver drowsiness estimation using an online multi-view and transfer TSK fuzzy system,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 3, pp. 1752–1764, 2021.

## Research Article

# ACEA: A Queueing Model-Based Elastic Scaling Algorithm for Container Cluster

Kui Li <sup>1,2</sup> Yi-mu Ji <sup>1,3,4,5</sup> Shang-dong Liu,<sup>1</sup> Hai-chang Yao,<sup>1</sup> Hang Li,<sup>1</sup> Shuai You,<sup>1</sup> and Si-si Shao<sup>1</sup>

<sup>1</sup>School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

<sup>2</sup>The Second People's Hospital of Nantong, Nantong 226002, China

<sup>3</sup>Nanjing Center of HPC China, Nanjing 210023, China

<sup>4</sup>Institute of High Performance Computing and Bigdata, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

<sup>5</sup>Jiangsu HPC and Intelligent Processing Engineer Research Center, Nanjing 210003, China

Correspondence should be addressed to Yi-mu Ji; [jiym@njupt.edu.cn](mailto:jiym@njupt.edu.cn)

Received 17 December 2020; Revised 19 January 2021; Accepted 10 May 2021; Published 26 May 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Kui Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Elastic scaling is one of the techniques to deal with the sudden change of the number of tasks and the long average waiting time of tasks in the container cluster. The unreasonable resource supply may lead to the low comprehensive resource utilization rate of the cluster. Therefore, balancing the relationship between the average waiting time of tasks and the comprehensive resource utilization rate of the cluster based on the number of tasks is the key to elastic scaling. In this paper, an adaptive scaling algorithm based on the queueing model called ACEA is proposed. This algorithm uses the hybrid multiserver queueing model ( $M/M/s/K$ ) to quantitatively describe the relationship among number of tasks, average waiting time of tasks, and comprehensive resource utilization rate of cluster and builds the cluster performance model, evaluation function, and quality of service (QoS) constraints. Particle swarm optimization (PSO) is used to search feasible solution space determined by the constraint relation of ACEA quickly, so as to improve the dynamic optimization performance and convergence timeliness of ACEA. The experimental results show that the algorithm can ensure the comprehensive resource utilization rate of the cluster while the average waiting time of tasks meets the requirement.

## 1. Introduction

With the advent of the information age, information technology has been widely used in various fields of human life, such as medical big data analysis in the medical field [1]. And the resulting application services have also been growing explosively. To provide an environment for an effective service running environment, cloud computing platforms based on virtualization technology emerge. Cloud computing refers to applications and services running on a distributed network using virtualized resources [2–4]. Virtualization is used to build virtual hosts running different operating systems on the same physical machine, while applications and services run on these different virtual machines as needed [5, 6]. In the early days of cloud computing [7], it was common to

build cloud computing clusters based on traditional virtual machine clusters. As Docker container technology [8, 9] is maturing day by day, the way of cloud platform construction has gradually changed into the shaping of a Docker cluster through the integration of multiple Docker physical nodes [10]. Compared with the traditional virtualization architecture [11–14], the container has the characteristics of low resource consumption, fast startup speed, high deployment efficiency, and good scalability, which can ensure the reliability and timeliness of the elastic scaling of resources for the cluster. However, the following problems still exist in the elastic scaling:

- (1) The container cluster has the characteristics of large number of resource indicators and complex

relationship between indicators, so it is difficult to analyze the relationship between indicators quantitatively. At this time, if the performance model and evaluation function [15–17] cannot be reasonably built according to user demand, system resource consumption, and other indicators, it is easy to have unreasonable allocation of resources

- (2) The number of tasks in the Internet environment has the characteristics of mutation, that is, the number of tasks arriving is irregular and sudden. When the elastic scaling algorithm cannot allocate resources according to the number of tasks in time, the task may be lost due to the long average waiting time of tasks or the resource waste caused by the cluster idling due to the insufficient number of tasks

Therefore, it is urgent to solve the problem of how to quantitatively describe the relationship among number of tasks, average waiting time of tasks, and comprehensive resource utilization rate of cluster, for the purpose of ensuring that the comprehensive resource utilization rate of cluster is always at a high level on the basis of controllable average waiting time of tasks. In order to solve the problem, this paper proposes the ACEA. The main contributions of this algorithm are as follows:

- (1) According to the state information of the container cluster and the application characteristics it carries, building a self-defined QoS constraint relationship and feasible solution space to provides computing constraints for the resources elastic scaling
- (2) The paper studies the problem of resource elastic scaling in container cluster, introduces the calculation method of task effective arrival rate, and proposes an adaptive elastic capacity expansion framework and performance model based on a queuing model. This paper uses  $M/M/s/K$  to describe the relationship among the number of tasks, the average waiting time of tasks, and the comprehensive utilization of cluster resources and solves the problem of building the performance model and evaluation function of a container cluster
- (3) Taking the cluster evaluation function as the fitness function, through the particle swarm optimization algorithm to search the feasible solution space, improves the dynamic optimization and convergence timeliness of ACEA algorithm, and achieves the goal of improving the accuracy of resource elastic scaling under the condition of ensuring the convergence timeliness of the algorithm

The structure of this paper is as follows. Section 2 summarizes the research work on the elastic scaling of container cluster. Section 3 discusses the overall design of the ACEA algorithm. Section 4 discusses the  $M/M/s/K$  performance model design and adaptive scaling strategy of the ACEA algorithm. Section 5 validates the overall design of Section 3 and the performance model and cluster schedul-

ing strategy of Section 4 through experiments. Section 6 is the conclusions.

## 2. Related Work

At present, the research mainly provides appropriate resources for the task under the premise of ensuring the shortest task execution time [18–20]. For example, [18] proposes a cloud environment task scheduling algorithm based on the multipriority queue and memory algorithm (MPQMA). The basic idea of this method is to improve the convergence speed with the advantages of MA. [19] provides a comprehensive multiobjective optimization task scheduling model to minimize execution time, delivery time, and execution cost. However, the scheduling model has conflicts of objective function parameters, and there may be the issue of timeliness in multiobjective optimization. [20] proposes a method for automatically testing the entire cloud environment using containers, which serves as the foundation of distributed cloud monitoring. In [21], the cluster elastic scaling technique is divided into reaction scaling and prediction scaling. Reaction scaling refers to the dynamic scaling of a cluster when a burst task request occurs. Prediction scaling refers to predicting task size based on historical data and prediction algorithms and dynamically scale before the change takes place. [22] proposes a layer-by-layer elastic scaling technique for applications. [23] discusses the application of optimization algorithms in load balancing and elastic scheduling. [24] proposes the application of the integrated MOPSO algorithm in task scheduling and optimizes the total task time and average task time. [25–28] introduce three resource scheduling methods provided by Docker Swarm: spread strategy, binpack strategy, and random strategy. The spread strategy is the default strategy. Docker Swarm prefers nodes with the fewest resources (such as CPU and memory) to ensure uniform use of all node resources in the cluster. The binpack strategy is the opposite of spread strategy, and its purpose is to use one node as much as possible to ensure enough idle nodes; the random strategy is a random selection strategy, that is, the task is completely randomly assigned to the existing nodes. Based on the theoretical analysis, the corresponding advantages and disadvantages of the scheduling algorithm are discussed. Among them, [25, 26] introduce Docker Swarm, the most widely used Docker cluster management tool, and provide the spread strategy as the default scheduling strategy. Docker Swarm selects the least quantity of resources to consume according to the number of CPU cores of a node and the unallocated memory; [28] proposes a task scheduling technique based on the genetic algorithm, which effectively allocates cloud computing resources and minimizes the overall response time. [29–31] detail the container orchestration tool Kubernetes and its elastic scaling function. The elastic scaling of Kubernetes can dynamically adjust the number of Pod copies for purpose of scaling the container according to the number of tasks. The number of Pod copies is adjusted by periodically querying the status of the Pod to obtain the monitoring data of the Pod, and then, the average usage rate and target usage rate of the existing Pod is compared to determine the number of scaling  $s$ . The

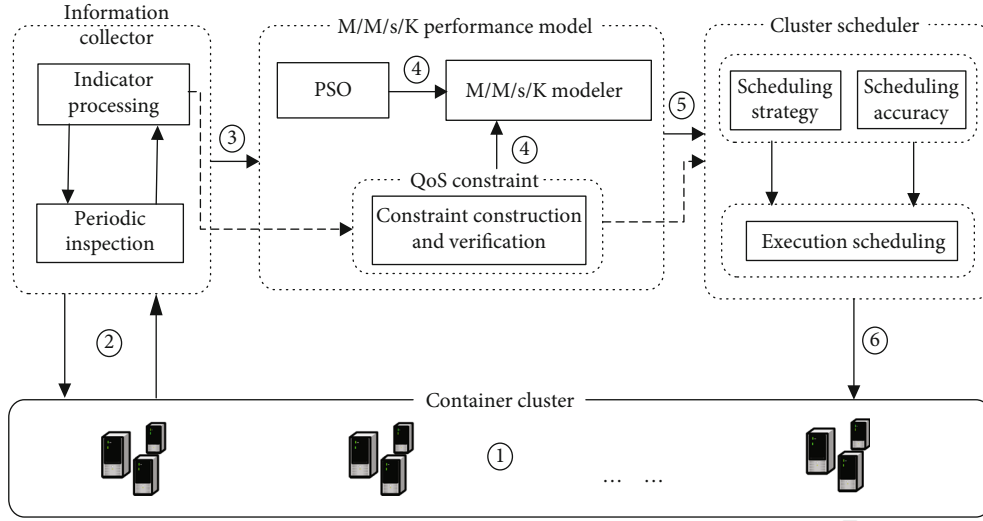


FIGURE 1: ACEA framework.

fuzzy system is famous for the good balance between approximation accuracy and interpretability. [32] uses the fuzzy system for data preprocessing to improve the accuracy of the algorithm. In this paper, whether we can use the above ideas for reference to classify and deal with tasks is found.

In summary, the existing container cluster dynamic scaling algorithm mainly provide resources for tasks under the premise of ensuring the minimum task execution time [18–20], without considering quantitatively the relationship among the sudden change of the number of tasks, average waiting time of tasks, and comprehensive resource utilization rate of cluster in the Internet environment.

### 3. Design of ACEA Algorithm

The overall design is shown in Figure 1. The algorithm ACEA consists of three modules: the information collector, M/M/s/K performance model, and cluster scheduler. The information collector is used to obtain the status of the current cluster and provide input data for M/M/s/K; the M/M/s/K performance model is the core of ACEA, which can be divided into three functional components: the QoS constraint verifier, M/M/s/K modeler, and PSO. The module mainly completes the construction of the cluster performance model, evaluation function and QoS constraints, verification of QoS constraints, and dynamic optimization and provides input data for the cluster scheduler. The cluster scheduler completes the specific cluster scheduling function according to the optimal number of containers, and the cluster state provided by the M/M/s/K performance model. The three modules of the ACEA algorithm are executed in sequence and form a closed loop.

#### 3.1. The Process of ACEA

- (1) The container cluster is responsible for receiving and processing tasks, which is the processing object of elastic scaling of the algorithm

- (2) The information collector is responsible for obtaining the current status of the cluster, such as CPU resources  $R_{\text{usedCPU}}$ , memory resources  $R_{\text{usedMEN}}$ , IO resources  $R_{\text{usedIO}}$ , network resources  $R_{\text{usedNET}}$ , and number of tasks  $n$ , etc. used by each container
- (3) After the M/M/s/K performance model obtains the data of step (2), firstly, the M/M/s/K modeler is used to quantitatively describe the relationship among the various indicators using the hybrid multiserver queuing model M/M/s/K and to build the performance model for the container cluster and the evaluation function which is also the fitness function of the PSO (see Section 3.2.1 for details)
- (4) Firstly, the QoS constraint verifier is used to construct QoS constraints. Then, under the constraints of the verifier, the cluster evaluation function in step (3) is used as the fitness function to solve the dynamic optimization problem and obtain the optimal number of containers in the current cluster and achieve the goal of dynamic optimization of the performance of the cluster (see Section 3.2.2 and Section 3.2.3 for details)
- (5) After the cluster scheduler obtains the optimal number of containers output by the M/M/s/K performance model module, the cluster scheduling strategy is determined according to the optimal numbers of containers and the state of the cluster.
- (6) Complete the cluster scheduling. After the scheduling is completed, go to step (2) to continue execution and form a closed loop.

**3.2. M/M/s/K Performance Model.** The M/M/s/K performance model is the core of ACEA. It consists of three functional components: the QoS constraint verifier, M/M/s/K modeler, and PSO. This section mainly discusses the components of the M/M/s/K performance model according to the overall design of the ACEA algorithm.



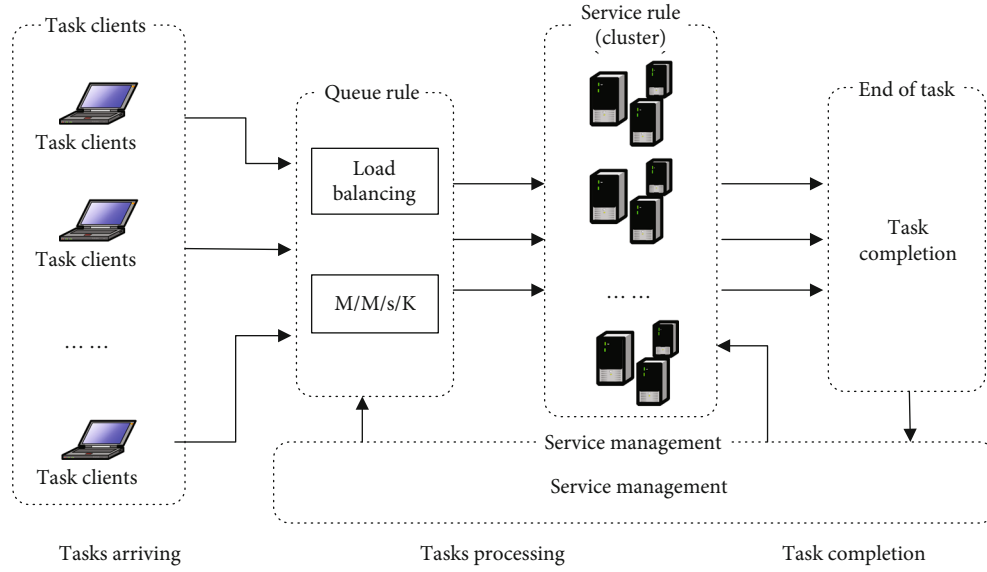


FIGURE 2: The model of the  $M/M/s/K$  modeler.

3.2.1. *M/M/s/K Modeler.* As shown in Figure 2, the ACEA algorithm compares tasks and containers providing services to customers and servers, respectively. When the container cluster processing capability cannot meet the QoS constraints, the number of containers in the cluster can be dynamically adjusted according to the number of tasks to enhance the service processing capability for the purpose of a flexible supply of resources.

The  $M/M/s/K$  modeler uses the principle of Figure 2 to quantitatively describe the relationships among the number of tasks, the average waiting time of tasks, and the comprehensive resource utilization rate of cluster using the hybrid multiserver queuing model  $M/M/s/K$ , so as to solve the construction problem of cluster performance model and evaluation function. The evaluation function is used as the fitness function of dynamic optimization. The reasons for choosing the queuing model  $M/M/s/K$  are as follows [33]:

- (1) The arrival of tasks and the processing time have relatively stable frequencies, while the task has discreteness and independence, which satisfies the condition of exponential distribution
- (2) There is an upper limit for the tasks that the servers can handle, which is consistent with the concept of “system space” in the queuing model. The upper limit is defined as  $K$
- (3) When the number of tasks to be processed in the servers reach  $K$ , the newly arrived task cannot be effectively processed in accordance with the requirements of the quality standard, which results in the loss of tasks. This is consistent with the principle of “when  $K$  locations have been occupied by customers, the newly arrived customers leave automatically” in the queuing model

- (4) When the number of tasks to be processed in the system is lower than  $K$ , the newly arrived task enters the queue and waits for the service, and the principle of “newly arrived customers enter the system to wait in line when the system has a free position” is consistent with the queuing model

3.2.2. *QoS Constraint Verifier.* The QoS constraint verifier mainly includes two functions: one is to construct QoS constraints and the other is to determine the solution space of the algorithm according to the QoS constraints. QoS constraints define the constraints among number of tasks, average waiting time, comprehensive resource utilization rate of cluster, and the cluster running indicators, which is the basis of constructing the QoS constraints verifier. They specifically include the following contents:

- (1) Maximum number of running containers: the maximum number of containers that the cluster hardware resources can support is denoted by  $K$ . In the production environment, the number of containers running in the cluster should be less than the maximum number of running containers. Otherwise, the container cannot be started due to insufficient hardware resources of the cluster, so there is  $s < K$
- (2) Average waiting time of tasks: the mathematical expectation of the maximum waiting time that the user can withstand from the time when the task is issued to the time when the cluster starts responding. If waiting time exceeds the average waiting time, the task will be lost. This paper assumes that the average waiting time of tasks does not exceed 50 ms, that is,  $W_q < 50$  ms
- (3) Queue length  $n$ : the number of tasks in the current queue. If  $s > n$ , which means the number of

containers is greater than the number of tasks, then there are free containers, resulting in waste of cluster resources. If  $n > K$ , which means the cluster is overloaded with tasks, it cannot process additional tasks effectively, resulting in the loss of these tasks; therefore, there must be  $s < n \leq K$

(4) Threshold constraint

Define  $f_{\text{used}i}$  as the weighted sum of resources consumed by the  $i$ th container in the presence of a task, and the following relationship exists.

$$f_{\text{used}i} = aR_{\text{usedCPU}}^i + bR_{\text{usedMEM}}^i + cR_{\text{usedIO}}^i + dR_{\text{usedNET}}^i. \quad (1)$$

Define  $f_{\text{total}i}$  as the weighted sum of the resources assigned to the  $i$ th container by the system, and the following relationship exists.

$$f_{\text{total}i} = aR_{\text{totalCPU}}^i + bR_{\text{totalMEM}}^i + cR_{\text{totalIO}}^i + dR_{\text{totalNET}}^i. \quad (2)$$

$R_{\text{usedCPU}}^i$  represents the CPU resource used by the  $i$ th container, and  $R_{\text{totalCPU}}^i$  indicates the total CPU resources allocated by the system for the  $i$ th container. Other resources are similar.  $a$ ,  $b$ ,  $c$ , and  $d$  are the weights of each resource in the total resources and are determined by the task attributes processed by the container, satisfying the relationship of  $a + b + c + d = 1$ ; this paper assumes  $a = b = c = d = 0.25$ .

Define  $U_i$  as the comprehensive resource utilization rate of a single container, derived from Equation (1), (2).

$$U_i = \frac{f_{\text{used}i}(R_{\text{usedCPU}}^i, R_{\text{usedMEM}}^i, R_{\text{usedIO}}^i, R_{\text{usedNET}}^i)}{f_{\text{total}i}(R_{\text{totalCPU}}^i, R_{\text{totalMEM}}^i, R_{\text{totalIO}}^i, R_{\text{totalNET}}^i)}. \quad (3)$$

$U_{\text{down}}$  is defined as the lower limit of the comprehensive resource utilization rate of cluster, which means that the comprehensive resource utilization rate of cluster is the lowest. If the utilization rate is lower than this value, it needs to shrink.  $U_{\text{up}}$  is defined as the upper limit of the comprehensive resource utilization rate of cluster, which means that the comprehensive resource utilization rate of cluster is the highest. If it is higher than this value, it needs to be scaling. According to the requirements for the comprehensive utilization of the cluster, the following constraint is obtained from Equation (3).

$$U_{\text{down}} < \sum_{i=1}^s \frac{U_i}{s} * 100\% < U_{\text{up}}. \quad (4)$$

**3.2.3. Particle Swarm Optimization.** Particle swarm optimization (PSO) algorithm makes use of an individual's sharing of information in the swarm, so that the swarm can evolve from disorder to order in the solution space to obtain the optimal solution. Due to its simple operation and fast convergence, PSO has been widely used in many fields such as function optimization, image processing, and geodetic survey [24]. The reasons for choosing PSO are as follows:

- (1) PSO has many mature applications in function optimization
- (2) PSO has a fast convergence rate and meets the timeliness requirements of the algorithm
- (3) PSO is easy to operate for improving computational efficiency

In the dynamic optimization solution, PSO uses the cluster evaluation function built by the M/M/s/K modeler in the solution space determined by the QoS constraint verifier to search for the optimal solution for the fitness function and obtain the obtained solution. The optimal solution is provided as input data to the cluster scheduler.

## 4. ACEA Performance Model and Scheduling Strategy Design

This section focuses on the design of the M/M/s/K performance model based on the overall design of ACEA and implements the processing flow and scaling conditions of each module through pseudocode.

**4.1. M/M/s/K Performance Model.** According to the existing cluster state and task attributes, ACEA firstly obtains the mathematical distribution parameters of task arrival and processing time by using statistical principle, then passes the obtained parameters into the M/M/s/K performance model, and finally solves the optimal expansion strategy under the constraints of QoS. The task processing flow is shown in Figure 3.

$L_q$  is defined as the task average queue length: the mathematical expectation of the number of tasks to be processed in the queued model. According to the Equations (1) and (3) constraints in Section 3.2.2, the following relationships exist:

$$L_q = \sum_{n=1}^K (n-s)p_n = \frac{\rho_0 \rho_s^s \rho_s}{s!(1-\rho_s)^2} [1 - \rho_s^{K-s+1} - (1-\rho_s)(K-s+1)\rho_s^{K-s}], \quad (5)$$

where  $\rho = \lambda/\mu$ ,  $\rho_s = \lambda/s\mu$  indicating service intensity, reflecting the busy degree of the system.

$$\rho_0 = \left( \sum_{n=0}^{s-1} \frac{\rho^n}{n!} + \frac{\rho^s (1 - \rho_s^{K-s+1})}{s!(1-\rho_s)} \right)^{-1}, \quad (6)$$

$$p_n = \frac{\rho^n}{s!s^{n-s}} \rho_0. \quad (7)$$

From Equations (5)–(7),  $W_q$  can be obtained.

$$W_q = \frac{L_q}{\lambda_e}. \quad (8)$$

$\lambda_e = \lambda(1 - p_k)$  indicates the effective arrival rate of tasks. The reason for the effective arrival rate of the task request is

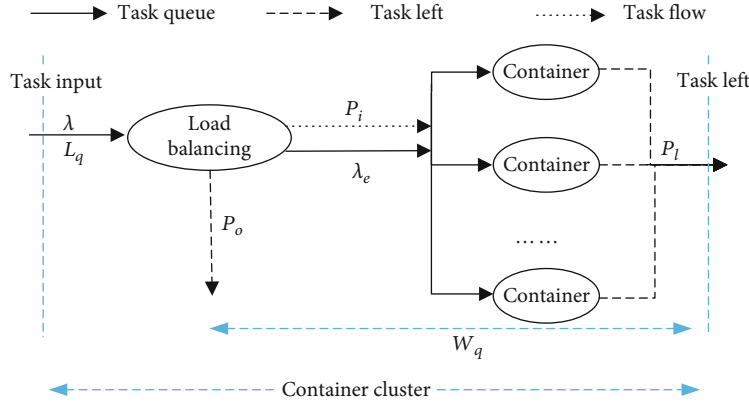


FIGURE 3: ACEA task processing flow.

that the part of the task failed to process properly during the cluster service. The reason for this is that in the process of cluster service, tasks with probability  $P_o$  cannot be handled properly, and tasks with probability  $P_i$  can be handled properly. Therefore, the arrangement Equation (8) can be obtained.

$$W_q = \frac{L_q}{\lambda_e} = \frac{\sum_{n=1}^K (n-s)p_n}{\lambda P_i} = \frac{\sum_{n=1}^K (n-s)p_n}{\lambda(1-P_o)}. \quad (9)$$

Because of the existence of QoS constraints, Equation (9) must satisfy the  $W_q < W_{q \max}$ , that is, there are the following relationships:

$$W_q = \frac{L_q}{\lambda_e} = \frac{\sum_{n=1}^K (n-s)p_n}{\lambda P_i} = \frac{\sum_{n=1}^K (n-s)p_n}{\lambda(1-P_o)} < W_{q \max}, \quad (10)$$

$$E[W_q | s = s] < W_{q \max} \text{ and } E[W_q | s = s - 1] \geq W_{q \max}, \quad (11)$$

$$U_{\text{down}} < E\left[\sum_{i=1}^s \frac{U_i}{s} \mid s = s\right] < U_{\text{up}}, \quad (12)$$

$$U_{\text{down}} > E\left[\sum_{i=1}^s \frac{U_i}{s} \mid s = s + 1\right] \text{ or } E\left[\sum_{i=1}^s \frac{U_i}{s} \mid s = s - 1\right] > U_{\text{up}}. \quad (13)$$

Equations (10)–(13) show that  $W_q$  is only related with  $s$ ,  $K$ ,  $\lambda$ , and  $\mu$ , and since  $\lambda$ ,  $\mu$ , and  $K$  are relatively independent,  $s$  that conforms to the QoS constraints Section 3.2.2 can be regarded as a set of feasible solutions, and all the set of feasible solutions is the solution space. Therefore, the problem of obtaining the optimal index of cluster turns into the problem of obtaining the best feasible solution. For this reason, PSO is introduced to search for the optimal solution in the solution space to achieve the goal of dynamic optimization.

**4.2. Adaptive Scaling Strategy Design.** This section implements the overall design and cluster scheduling strategy of ACEA in pseudocode. The pseudocode is shown in Algorithm 1. Line 1 defines the model parameters of the algorithm as global variables, including the distribution function parameters, the

number of particles, the maximum number of iterations, and the adaptation degree of the task arrival and processing in the queuing model M/M/s/K. Line 2 defines the threshold of the system's comprehensive resource utilization rate and average waiting time of tasks. Line 3 defines the configuration file function of information collection service, which is used to obtain the cluster's state parameters, including CPU usage, memory usage, IO resources, network resources, etc. Lines 4-5 construct the particle fitness function (performance model evaluation function) and particle initialization and particle swarm algorithm according to the parameters of Line 1. Lines 6-12 particles search the feasible solution space for dynamic optimization. Lines 14-16 define the condition for scaling, that is, if the average waiting time of tasks satisfies the requirement, and the fitness value is lower than the lower threshold, contraction() function is performed. Lines 17-19 define the condition for scaling, that is, if the average waiting time of tasks satisfies the requirement and the fitness value is higher than the upper threshold, expand() function is performed. Lines 20-22 defines the condition for stability, that is, if the average waiting time of tasks satisfies the requirement, and the fitness value is between the upper and lower thresholds, scheduling is not performed, and only the current results will be visually managed. There are three main scheduling strategies for Algorithm 1.

- (1) Container Contraction. The main reason for the container contraction is that the number of tasks is reduced or the task processing is completed, causing the decrease of various monitoring indicators to different extents. The decrease of  $\sum f_{\text{used } i}$  will make the resource utilization rate lower than the MIN threshold. In this case, the container cluster needs to be reduced
- (2) Container Expand. The main reason for the expansion of the container is that the number of tasks increases, causing the increase the monitoring indicators to different extents. The increase of  $\sum f_{\text{used } i}$  will make the resource utilization rate higher than the upper threshold or the average waiting time of tasks failing to meet the QoS constraints. In this case, the container cluster needs to be expanded

```

Input: achievement rate of historical tasks  $\lambda_{\text{History}}$ , task processing interval  $\mu$ , number of containers in the current cluster  $s$ .
Output: indicator information for container cluster, volume of autoscaling, mode of adaptive scaling.
1. global var  $W_q$ , Fitness; //Fitness: real-time comprehensive resource utilization rate of cluster.
2. const var  $U_{\text{up}}$ ,  $U_{\text{down}}$ ,  $W_{\text{limit}}$ ,  $U_{\text{down}}$ ; //  $W_{\text{limit}}$ : maximum waiting time of tasks.
3. define func getConfig():  $L_{\text{CPU}}$ ,  $L_{\text{MEM}}$ ,  $L_{\text{IO}}$ ,  $L_{\text{NET}}$ ,  $L_{\text{LIMIT}}$ ;
4. Input
5. par[Pnum]=
6.   PSOInit(Pnum): ParticleInit(Pnum), ParticleEvaluate();
7.   Fitness = Wq(getUall(s), getUlimit());
8. PSORun():
9.   For each par.
10.    ParticleUpdate();
11.    ParticleEvaluate();
12.   End For;
13. For Iteration times do.
14.   If ( $W_q < W_{\text{limit}}$  &&  $U_{\text{down}} > \text{Fitness}$ ):
15.     Contract(service);
16.   End if;
17.   If( $(W_q < W_{\text{limit}}$  &&  $\text{Fitness} > U_{\text{up}})$  ||  $W_q < W_{\text{limit}}$ ):
18.     Expand(service);
19.   End if;
20.   If( $W_q < W_{\text{limit}}$  &&  $U_{\text{down}} < \text{Fitness} < U_{\text{up}}$ ):
21.     PSOshowresult();
22.   End if;
23. End For;

```

ALGORITHM 1: The algorithm of ACEA.

- (3) Stable State. The main reason for the stability of the cluster is that the number of tasks is stable and there is no sudden change, so that the average waiting time of tasks and resource utilization rate are in line with the custom QoS constraints

usage rate of the existing Pod with the target usage rate to determine the number of Pod copies, and finally through the horizontal dynamic adjustment of the number of Pod copies to achieve the purpose of scaling. The formula for calculating the elastic scaling of Pod is as follows.

## 5. Results and Discussion

The algorithm ACEA has been prototyped in the Docker virtualized cluster. This section will verify the effectiveness of the algorithm in the case of a sudden change in the number of tasks and compare with the existing elastic scaling algorithm to verify the accuracy of the system's model. The objects for comparison are two general algorithms in the field of elastic scaling.

The incremental scheduling algorithm (ISA) is an algorithm that periodically checks the state of the cluster through the polling service. When the task average waiting time or the comprehensive resource utilization rate of cluster does not meet the QoS constraints, the cluster scheduler or operation and maintenance personnel, based on historical experience, will quantitatively determine the number of containers to be adjusted in a certain interval to cope with the current task. The quantitative determination of the number of containers required to be adjusted within a certain interval is an increment.

Kubernetes HPA [29] (Kubernetes horizontal Pod autoscaling, Kubernetes Pod) obtains information about resource usage by periodically polling the Pod state during the operation of the container cluster and then compares the average

$$\text{ExpansionPods} = \frac{\text{Ceil}(\text{Sum}(\text{CurrentUtilization}))}{\text{TargetUtilization}}. \quad (14)$$

Among them, ExpansionPods indicates the number of containers required; Target Utilization indicates the user-defined resource usage threshold; CurrentUtilization indicates the average resource utilization of the current Pod, and the calculation formula is as shown in Equation (15); Sum() is a summation function for the sum of current utilization; Ceil() is the integer function used to return the smallest integer greater than or equal to the specified expression.

$$\text{CurrentUtilization} = \frac{\text{Average value of used resources}}{\text{Resources allocated to Pod by the system}}. \quad (15)$$

*5.1. Experimental Threshold.* The experimental threshold settings in this section are as follows:

- (1) The upper threshold of the average waiting time of tasks is set to 50 ms, that is  $W_q \leq 50$  ms
- (2) The upper threshold of the comprehensive resource utilization rate of cluster is 80%, and the lower

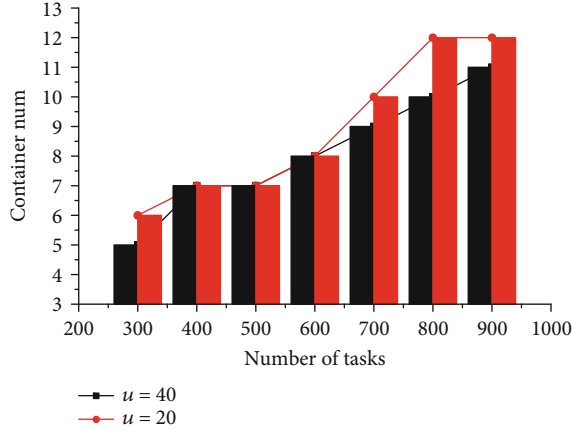


FIGURE 4: Adaptive figure of the number of containers.

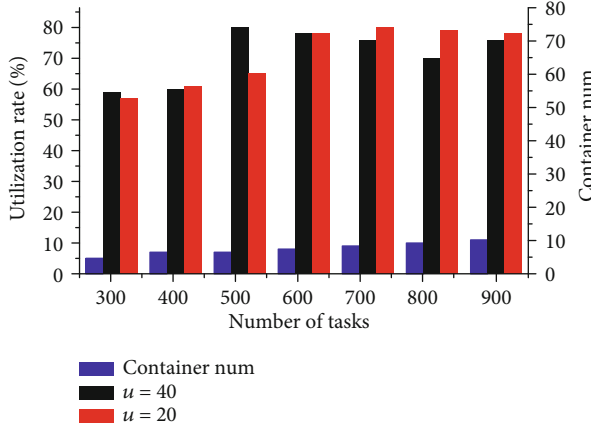


FIGURE 5: Changing figure of the comprehensive resource utilization rate.

threshold is 55%. Equation (4) gives the relationship of cluster resource utilization as shown in Equation (16)

$$55\% < \sum_{i=1}^s \frac{f_{\text{used}i} (R_{\text{usedCPU}}^i, R_{\text{usedMEM}}^i, R_{\text{usedIO}}^i, R_{\text{usedNET}}^i)}{f_{\text{total}i} (R_{\text{totalCPU}}^i, R_{\text{totalMEM}}^i, R_{\text{totalIO}}^i, R_{\text{totalNET}}^i)} * 100 < 80\%. \quad (16)$$

**5.2. Experimental Analysis.** The purpose of this experiment is to verify the coordination and effectiveness of the information collector, the M/M/s/K performance model, cluster scheduler in ACEA, and the feasibility of ACEA's scaling strategy. The experimental results are shown in Figures 4–6.

As shown in Figure 4, the number of containers in the cluster changes simultaneously when the number of tasks changes, and the trend of change is consistent with the number of tasks, which verifies the feasibility of container cluster scheduling strategy. The following validation experiments are based on the experimental data to verify the performance of the average waiting time of tasks and the comprehensive resource utilization rate of cluster.

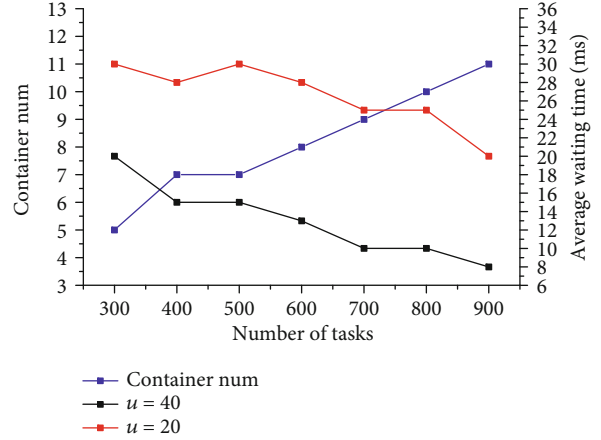


FIGURE 6: Changing figure of the average waiting time of tasks.

As shown in Figure 5, when the number of tasks changes, the number of containers in the cluster and the comprehensive resource utilization rate of cluster change accordingly. Corresponding to the left axis of the figure above, the number of tasks and the number of containers have the same trend. Corresponding to the right axis of the figure above, although the number of tasks that the cluster can handle per unit time, i.e., the average service rate  $\mu$ , is different, the comprehensive utilization rate of cluster resources calculated by Equations (3) and (4) is always between 55% and 80%, which meets the threshold requirement of the comprehensive resource utilization rate of cluster.

As shown in Figure 6, when the number of tasks changes, the number of containers in the cluster and the average waiting time of tasks change accordingly. Corresponding to the left axis of the figure above, the number of tasks and the number of containers have the same trend. Corresponding to the right axis of the figure above, although the number of tasks that the cluster can handle per unit time, i.e., the average service rate  $\mu$ , is different, the average waiting time of tasks increases synchronously and is always lower than 35 ms, which satisfies the threshold requirement of average waiting time of tasks and does not affect the effective processing of tasks.

It can be seen from the above experimental results that the average waiting time of tasks is always within 35 ms, and the comprehensive resource utilization rate of cluster is always maintained between 55% and 80%, which satisfies the requirement for the threshold set by the user, indicating that the algorithm satisfies the average waiting time of tasks. In the case of time requirements, the comprehensive resource utilization rate of cluster is also guaranteed.

**5.3. Comparison with ISA Algorithm.** The biggest feature of this algorithm is easy to implement, no need for plugins, and easy operation and maintenance personnel. Compared with the algorithm ACEA, the performance is shown in Figures 7 and 8.

As shown in Figure 7, with the running of the system, when the number of tasks changes, the number of containers in the cluster and the average waiting time of tasks change accordingly. Corresponding to the left axis of the figure

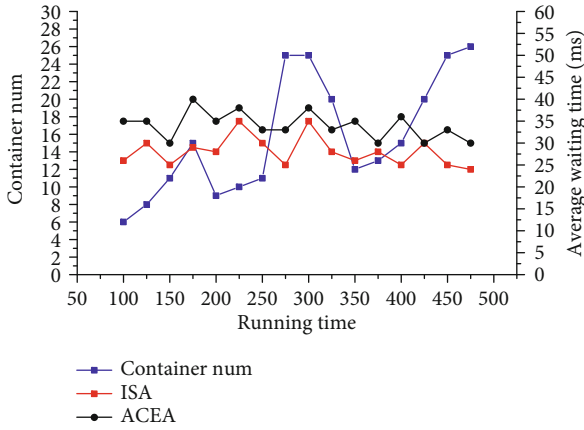


FIGURE 7: Comparison of average waiting time of tasks.

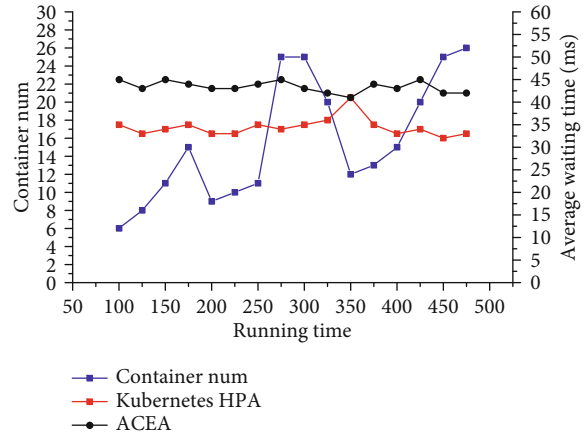


FIGURE 9: Comparison of average waiting time of tasks.

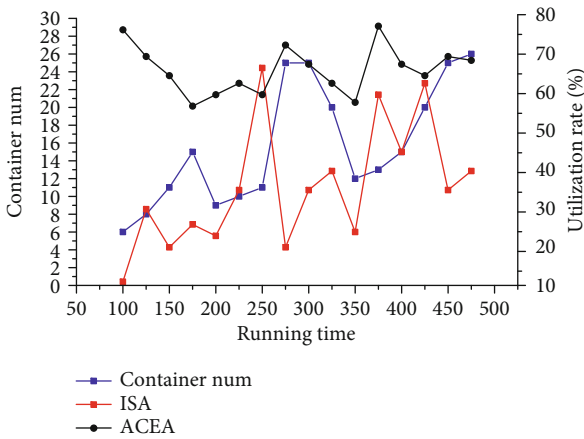


FIGURE 8: Comparison of the comprehensive resource utilization rate.

above, with the system running, in order to meet the threshold requirements, the number of containers in the cluster changes when the number of tasks changes. Corresponding to the right axis of the figure above, the average waiting time of ACEA algorithm is higher than that of ISA algorithm, and the difference is less than 12 ms. Meanwhile, although the average waiting time of ACEA algorithm fluctuates, it is always lower than 45 ms, which satisfies the threshold requirement of average waiting time of tasks and does not affect the effective processing of tasks.

As shown in Figure 8, with the running of the system, when the number of tasks changes randomly, the comprehensive resource utilization rate and the number of containers in the cluster will change accordingly. For effective comparison of the performance of the algorithm, the change rule of the number of tasks here is consistent with Figure 7. Corresponding to the left axis of the figure above, with the running of the system, in order to meet the threshold requirements, the number of containers in the cluster changes when the number of tasks changes. Corresponding to the right axis of the figure above, when the running time is 250, the comprehensive resource utilization rate of ISA algorithm is higher than that of ACEA algorithm. The reason is that the number of capacity expansion determined by experience,

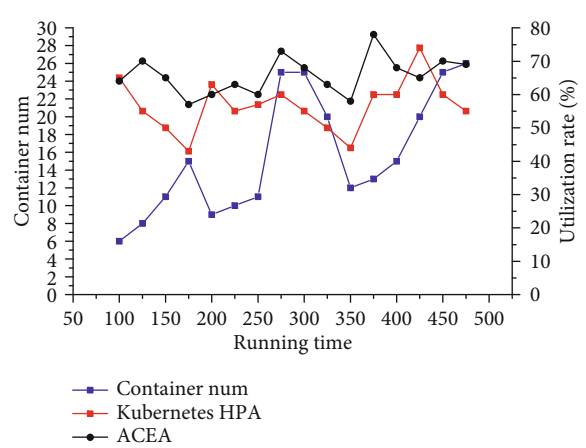


FIGURE 10: Comparison of the comprehensive resource utilization rate.

i.e. the increment, matches the number of tasks at present, but overall, the comprehensive resource utilization rate of ACEA algorithm is higher than that of ISA algorithm, and the comprehensive resource utilization rate of ACEA algorithm is always between 55% and 80%, which achieves the goal of ACEA algorithm to improve the comprehensive resource utilization rate on the basis of ensuring the average waiting time of tasks.

5.4. Comparison with Kubernetes HPA. From the perspective of theoretical analysis, Kubernetes HPA does not provide the function of automatically scaling the Pod according to the constraint of average waiting time of tasks. ACEA makes up for this part of the blank and ensures the balance between the average waiting time of tasks and the comprehensive resource utilization rate of cluster by using the dynamic optimization algorithm. From an experimental point of view, the performance is shown in Figures 9 and 10 compared to ACEA.

As shown in Figure 9, with the running of the system, when the number of tasks changes, the number of containers in the cluster and the average waiting time of tasks change accordingly. Corresponding to the left axis of the figure

above, with the system running, in order to meet the threshold requirements, the number of containers in the cluster changes when the number of tasks changes. Corresponding to the right axis of the figure above, the average waiting time of ACEA algorithm is higher than that of Kubernetes HPA, and the difference is less than 15 ms. Meanwhile, although the average waiting time of ACEA algorithm fluctuates, it is always lower than 45 ms, which satisfies the threshold requirement of average waiting time of tasks and does not affect the effective processing of tasks.

As shown in Figure 10, with the running of the system, when the number of tasks changes randomly, the comprehensive resource utilization rate and the number of containers in the cluster will change accordingly. For effective comparison of the performance of the algorithm, the change rule of the number of tasks here is consistent with Figure 9. Corresponding to the left axis of the figure above, with the running of the system, in order to meet the threshold requirements, the number of containers in the cluster changes when the number of tasks changes. Corresponding to the right axis of the figure above, when the running time is 200 and 425, the comprehensive resource utilization rate of Kubernetes HPA is higher than that of ACEA algorithm. The reason is due to the influence of Kubernetes HPA resource monitoring transmission delay and Equation (10) average strategy, but overall, the comprehensive resource utilization rate of ACEA algorithm is higher than that of Kubernetes HPA, and the comprehensive resource utilization rate of ACEA algorithm is always between 55% and 80%, which achieves the goal of ACEA algorithm to improve the comprehensive resource utilization rate on the basis of ensuring the average waiting time of tasks.

## 6. Conclusions

This paper proposes the ACEA algorithm for balancing the average waiting time of tasks and the comprehensive resource utilization rate of cluster in elastic scaling. This algorithm uses a hybrid multiserver queuing model  $M/M/s/K$  to build a container cluster performance model and evaluate functions and QoS constraints and uses a QoS constraint validator to determine the feasible solution space of the algorithm, and the feasible solution space is searched by PSO to achieve dynamic optimization of the algorithm. The experimental results show that the proposed algorithm can ensure the comprehensive resource utilization rate of cluster while guaranteeing that the average waiting time of tasks is satisfied. However, we believe that there is still some follow-up work worth extending, including the following:

- (1) For the characteristics of sudden changes of the number of tasks, it may be considered to introduce a multiple QoS authentication strategy or a smoothing algorithm to avoid invalid scaling and reduce the jitter of cluster scaling
- (2) Try to apply ACEA to the microservice architecture container cluster

- (3) At present, there are some improved clustering algorithms. We can try to cluster tasks on the basis of the improved clustering algorithm and use the task category attribute to improve the efficiency of task forwarding in load balancing

## Data Availability

The data used to support the findings of this study are included in the article.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Authors' Contributions

Kui Li did the conceptualization, methodology, writing—original draft, and writing—review and editing. Yi-mu Ji did the funding acquisition and project administration. Shang-dong Liu found resources and did the validation. Hai-chang Yao did the visualization. Hang Li and Shuai You acquired the software. Si-si Shao did the data curation and investigation.

## Acknowledgments

This work was supported by the National Key R&D Program of China (2020YFB2104000, 2020YFB2104002), Natural Science Foundation of Jiangsu Province (Higher Education Institutions) (BK20170900, 19KJB520046, and 20KJA520001), Innovative and Entrepreneurial talents projects of Jiangsu Province, Jiangsu Planned Projects for Postdoctoral Research Funds (No. 2019K024), Six talent peak projects in Jiangsu Province (JY02), Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYCX19\_0921, KYCX19\_0906), Zhejiang Lab (2021KF0AB05), and NUPT DingShan Scholar Project and NUPTSF (NY219132).

## References

- [1] L. Wang, Y. Zhang, M. Jiang et al., "Toward a normalized clinical drug knowledge base in China - applying the RxNorm model to Chinese clinical drugs," *Journal of the American Medical Informatics Association*, vol. 25, no. 7, pp. 809–818, 2018.
- [2] Z. H. Zhan, X. F. Liu, Y. J. Gong, J. Zhang, H. S. H. Chung, and Y. Li, "Cloud computing resource scheduling and a survey of its evolutionary approaches," *ACM Computing Surveys*, vol. 47, no. 4, pp. 1–33, 2015.
- [3] A. R. Hummada, N. W. Paton, and R. Sakellariou, "Adaptation in cloud resource configuration: a survey," *Journal of Cloud Computing*, vol. 5, no. 7, article 7, 2016.
- [4] J. V. B. Bibal and D. Dejeu, "An auto-scaling framework for heterogeneous Hadoop systems," *International Journal of Cooperative Information Systems*, vol. 26, no. 4, article 1750004, 2017.
- [5] M. Abdullah, K. Lu, P. Wieder, and R. Yahyapour, "A heuristic-based approach for dynamic VMs consolidation in cloud data centers," *Arabian Journal for Science & Engineering*, vol. 42, no. 8, pp. 3535–3549, 2017.

- [6] Y. T. Lin, M. L. Wen, M. Jou, and D. W. Wu, "A cloud-based learning environment for developing student reflection abilities," *Computers in Human Behavior*, vol. 32, pp. 244–252, 2014.
- [7] Z. Kozhირbayev and R. O. Sinnott, "A performance comparison of container-based technologies for the cloud," *Future Generation Computer Systems*, vol. 68, pp. 175–182, 2017.
- [8] "Docker—build, ship, and run any app, anywhere," January 2019, <http://www.docker.com>.
- [9] D. Merkel, "Docker: lightweight Linux containers for consistent development and deployment," *Linux Journal*, vol. 239, 2014.
- [10] F. Paraiso, S. Challita, Y. Al-Dhuraibi, and P. Merle, "Model-driven management of docker containers," in *2016 IEEE 9th International Conference on Cloud Computing (CLOUD)*, pp. 718–725, San Francisco, CA, USA, 2016.
- [11] Q. Wu, "Making Facebook's software infrastructure more energy efficient with autoscale," *Facebook Engineering Blog*, 2014, <https://engineering.fb.com/2014/08/08/production-engineering/making-facebook-s-software-infrastructure-more-energy-efficient-with-autoscale>.
- [12] A. M. Joy, "Performance comparison between linux containers and virtual machines," in *2015 International Conference on Advances in Computer Engineering and Applications*, pp. 342–346, Ghaziabad, India, 2015.
- [13] Z. Li, M. Kihl, Q. Lu, and J. A. Andersson, "Performance overhead comparison between hypervisor and container based virtualization," in *2017 IEEE 31st International Conference on Advanced Information Networking and Applications (AINA)*, pp. 955–962, Taipei, Taiwan, 2017.
- [14] J. A. Aroca, A. F. Anta, M. A. Mosteiro, C. Thraves, and L. Wang, "Power-efficient assignment of virtual machines to physical machines," *Future Generation Computer Systems*, vol. 54, pp. 82–94, 2016.
- [15] P. R. Desai, "A survey of performance comparison between virtual machines and containers," *International Journal of Computer Sciences and Engineering*, vol. 4, no. 7, pp. 55–59, 2016.
- [16] Y. Kouki and T. Ledoux, "SCALING: SLA-driven cloud auto-scaling," in *Proceedings of the 28th Annual ACM Symposium on Applied Computing - SAC '13*, pp. 411–414, Coimbra Portugal, 2013.
- [17] C. Liu, B. T. Loo, and Y. Mao, "Declarative automated cloud resource orchestration," in *Proceedings of the 2nd ACM Symposium on Cloud Computing - SOCC '11*, Cascais, Portugal, 2011.
- [18] B. Keshanchi and N. J. Navimipour, "Priority-based task scheduling in the cloud systems using a memetic algorithm," *Journal of Circuits, Systems and Computers*, vol. 25, no. 10, article 1650119, 2016.
- [19] F. Ramezani, J. Lu, and F. Hussain, "Task scheduling optimization in cloud computing applying multi-objective particle swarm optimization," in *Service-Oriented Computing. ICSOC 2013*, S. Basu, C. Pautasso, L. Zhang, and X. Fu, Eds., vol. 8274 of Lecture Notes in Computer Science, pp. 237–251, Springer, Berlin, Heidelberg, 2013.
- [20] S. Dhakate and A. Godbole, "Distributed cloud monitoring using Docker as next generation container virtualization technology," in *2015 Annual IEEE India Conference (INDICON)*, pp. 1–5, New Delhi, India, 2015.
- [21] D. Wang, S. Zhu, T. Li, and Y. H. Gong, "Comparative document summarization via discriminative sentence selection," *ACM Transactions on Knowledge Discovery from Data*, vol. 6, no. 3, pp. 1–18, 2012.
- [22] K. Lerman and R. McDonald, "Contrastive summarization: an experiment with consumer reviews," in *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Short Papers on - NAACL '09*, pp. 113–116, Boulder, Colorado, 2009.
- [23] X. Wan, H. Jia, S. Huang, and J. G. Xiao, "Summarizing the differences in multilingual news," in *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information - SIGIR '11*, pp. 735–744, Beijing, China, 2011.
- [24] M. Abdullah, E. A. Al-Muta'a, and M. A. Sanabani, "Integrated MOPSO algorithms for task scheduling in cloud computing," *Journal of Intelligent & Fuzzy Systems*, vol. 36, no. 2, pp. 1823–1836, 2019.
- [25] N. Naik, "Applying computational intelligence for enhancing the dependability of multi-cloud systems using docker swarm," in *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 1–7, Athens, Greece, 2016.
- [26] N. Naik, "Building a virtual system of systems using Docker Swarm in multiple clouds," in *2016 IEEE International Symposium on Systems Engineering (ISSE)*, pp. 1–3, Edinburgh, UK, 2016.
- [27] J. W. Xu, W. B. Zhang, T. Wang, and T. Huang, "A genetic algorithm based adaptive strategy for image backup of virtual machines," *Chinese Journal of Computers*, vol. 39, no. 2, pp. 351–362, 2016.
- [28] M. Agarwal and G. M. S. Srivastava, "A genetic algorithm inspired task scheduling in cloud computing," in *2016 International Conference on Computing, Communication and Automation (ICCCA)*, pp. 364–367, Greater Noida, India, 2016.
- [29] "Kubernetes plugin," January 2019, <http://docs.getcloudify.org/4.1.0/plugins/kubernetes>.
- [30] J. Opara-Martins, R. Sahandi, and F. Tian, "Critical analysis of vendor lock-in and its impact on cloud computing migration: a business perspective," *Journal of Cloud Computing*, vol. 5, article 4, 2016.
- [31] L. Nikolaos, "D1.1 requirements analysis report. Cloud 4SOA project deliverable," January 2019, <https://pdfs.semanticscholar.org/20fb/57b26982a404138a32ff756e73d26c29a6f2.pdf>.
- [32] Y. Zhang, F. L. Chung, and S. Wang, "Takagi-Sugeno-Kang fuzzy systems with dynamic rule weights," *Journal of Intelligent & Fuzzy Systems*, vol. 37, no. 6, pp. 8535–8550, 2019.
- [33] C. de Alfonso, A. Calatrava, and G. Moltó, "Container-based virtual elastic clusters," *Journal of Systems and Software*, vol. 127, pp. 1–11, 2017.



## Research Article

# Social Network Big Data Hierarchical High-Quality Node Mining

Dongning Jia <sup>1,2</sup>, Bo Yin <sup>1,2</sup> and Xianqing Huang <sup>2</sup>

<sup>1</sup>Ocean University of China, Qingdao Shandong 266100, China

<sup>2</sup>Pilot National Laboratory for Marine Science and Technology (Qingdao), Qingdao Shandong 266237, China

Correspondence should be addressed to Bo Yin; ybfirst@ouc.edu.cn

Received 9 April 2021; Revised 26 April 2021; Accepted 7 May 2021; Published 18 May 2021

Academic Editor: Shan Zhong

Copyright © 2021 Dongning Jia et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Compared with the conventional network data analysis, the data analysis based on social network has a very clear object of analysis, various forms of analysis, and more methods and contents of analysis. If the conventional analysis methods are applied to social network data analysis, we will find that the analysis results do not reach our expected results. The results of the above studies are usually based on statistical methods and machine learning methods, but some systems use other methods, such as self-organizing self-learning mechanisms and concept retrieval. With regard to the current data analysis methods, data models, and social network data, this paper conducts a series of researches from data acquisition, data cleaning and processing, data model application and optimization of the model in the process of application, and how the formed data analysis results can be used for managers to make decisions. In this paper, the number of customer evaluations, the time of evaluation, the frequency of evaluation, and the score of evaluation are clustered and analyzed, and finally, the results obtained by the two clustering methods applied in the analysis process are compared to build a customer grading system. The analysis results can be used to maintain the current Amazon purchase customers in a hierarchical manner, and the most valuable customers need to be given key attention, combining social network big data with micro marketing to improve Amazon's sales performance and influence, developing from the original single shopping mall model to a comprehensive e-commerce platform, and cultivating their own customer base.

## 1. Introduction

With the rapid development of computer technology, the Internet electronic information resources play an indispensable role in everyone's life and work, and people interact with each other more and more through the Internet [1]. Social networking is a change from online social networking, the predecessor of online social networking is email, and online social networking pushes BBS one step forward [2]. Email and BBS are upgraded versions of social tools are instant messaging and blogs, which have significantly improved in terms of transmission speed and parallel processing; blogs have begun to reflect sociological and psychological theories. In recent years, the development of social networks has been remarkable [3–5]. At present, more than half of Chinese Internet users communicate and share information through social networks, which have become the Web 2.0 business with the largest communication impact, the widest coverage of users, and the highest commercial value. Social networks

have become a part of people's lives and play a very important role in people's lives, having an undervalued impact on people's access to information, thinking, and life [6–8].

Social networks have become a window for people to get information, show themselves, and market and promote their functions. As of December 2012, the total number of Internet users is 564 million, and 309 million are microblog users, in which tens of millions of Internet users are active every day, posting microblogs to share what is new, etc. [9, 10]. Social networks have three major advantages: (1) higher user viscosity. According to the survey data, an average Internet user spends about 17 to 20 minutes a day on social networks. (2) Low maintenance cost. The number of editors needed for Web 1.0 portal websites exceeds the number of employees for Web 2.0 websites. (3) Information preparation. Social network users are asked to fill in real, detailed personal information. This facilitates developers to conduct data analysis as well as business applications. Web data is growing rapidly, and having access to this data on websites

can be of great help in analyzing topical social issues and trends and also can be of great help and impact on social network operations [11]. Social networks concentrate the youngest and most active web users, so they are also the most densely populated places for online speech, and mining their information resources can obtain a lot of valuable raw data for relevant user analysis and decision-making [12]. Data mining is a thriving disciplinary frontier concerning data and information systems and applications, the result of a multidisciplinary field and drawing from several disciplines, and a natural evolution from information technology [13–15]. It is able to precisely mine data hidden knowledge from massive amounts of data, and data mining is applied in any type of information repository and transient data. In data mining applications, the most basic forms of data are database data, data warehouse data, and transactional data [16, 17]. Data mining has now become one of the most cutting-edge research directions in the field of information decision-making and databases internationally, mainly because of its great potential for business prospects and has attracted wide attention from industry and academia [18, 19]. Data mining can be applied to financial data analysis, retailing, telecommunications, biological data analysis, and other scientific applications. In this paper, data mining is applied to data mining in social networks. The huge amount of data in the Internet and the data will grow rapidly, provides a good basis for data mining [20, 21].

In recent years, due to the limitations of search engine query information, many scholars improve the efficiency of retrieval is by eliminating the near mirror pages, the research of algorithms for near text detection. Researchers at the University of Arizona, USA, found similar documents existing in a large file system by employing a method that computes the degree of overlap of documents. The results of the studies presented above are generally based on statistical methods and machine learning methods, but some systems use other methods such as self-organizing self-learning mechanisms and concept retrieval. The diversity of web data forms brings new challenges to data mining, and it is the main effort and development direction in the future to combine clustering, support vector machines, neural networks, etc. with various database techniques in its large amount of graphical and complex spatiotemporal data, and to investigate data mining in new databases. In this paper, we study the application of data mining in social networks. The social network object studied in this paper is Sina Weibo. The data of popular topics and participating users in Sina Weibo are extracted by python, and then, the clustering algorithm and collaborative filtering algorithm of data mining are used to analyze and recommend the user data. The main work done is as follows.

- (1) It briefly introduces the basis of web data extraction and methods of extracting data, then introduces data mining techniques and clustering algorithms and hierarchical methods in clustering methods
- (2) This article describes in detail the extraction of user data from social networks in Python, using two

methods: the Weibo API interface and the simulated browser login method

- (3) Clustering of the extracted trending topics and topic recommendations to users based on their participation in the topics

## 2. Data Mining Techniques and Tools

*2.1. Data Mining Tools.* The Oracle database management system is used as the tool for data storage, which is a relational database and one of the current mainstream database management systems. Compared with other database management systems, it is easy to install, easy to use, and easy to understand interface. The simple and easy-to-use SQL statements are not necessary for beginners to consider how to process data with complex algorithms, and they can get the desired data with suitable SQL statements. Currently, there are two main categories of popular data mining tools: one is a data mining tool specifically for a specific industry or field, and the other is a general mining tool with a wide range of applications and more situations. Data mining tools for a specific field or industry, the tool itself has some optimization or preprocessing work on the data. In this paper, we choose the more general data mining tools, and the main general and free data mining tools are WEKA from New Zealand, R language with powerful plotting performance and statistical analysis, and the Clementine system from SPSS.

Among these three data mining tools, Clementine is the industry's leading data mining platform that can apply complex data mining algorithms and machine learning techniques to process data to help companies uncover the value behind transaction data, with the strongest ease of use and the most beautiful interface. The main algorithms applied in this data mining process are K-Means clustering algorithm, two-step clustering algorithm, RFM structural model, etc. The analysts using Clementine do not need to spend much time on the algorithms themselves, but only need to apply a reasonable data model to observe whether the data analysis results are consistent with the business needs.

### 2.2. Clementine Has Three Main Features

- (1) *Beautiful Interface and Visualization of Data Analysis Process.* As a general-purpose data mining tool, the beautiful and visualized operation interface is one of its great advantages. Users only need to select the nodes they want in the modules of source, record option, field option, graph, modeling, output, and export and connect several nodes with lines to complete the creation of a basic model.
- (2) *Powerful Data Processing Capability.* Even a novice who has no contact with data mining algorithms can use Clementine to build data models by reviewing some theoretical knowledge of algorithms and knowing the application scenarios of algorithms. The simple model in Clementine does not require any parameter setting; just input the preprocessed

data in the data source and then run and output the results. The expert model requires the analysts to select the established data model according to the business requirements and customize some model analysis parameters according to the actual needs in the analysis process. The setting of parameters mainly relies on the experience value of business personnel in the practice process to set and through the adjustment of parameters to continuously optimize the data model and build a model that meets the actual needs of the enterprise, in order to achieve the purpose of outputting ideal analysis results, providing data support for the development of business and providing a source of power for the advancement of the enterprise.

- (3) *Follow the Standard Data Mining Process of CRISP.DM.* Unlike the traditional mining process built on the technical level, Clementine can effectively control the entire mining process, equating the data mining process with a business analysis process and making business purposes and business needs the latest goal of data mining. A complete data analysis process includes six stages: business understanding, data understanding, data preparation, modeling, evaluation, and deployment, and different stages correspond to different project management requirements.

Following the CRISP.DM process enhances the user's understanding of the business model and achieves a perfect alignment between business requirements and the data model. More than half of the data mining platforms in the industry follow the CRISP.DM standard, which has gradually become the industry standard. As shown in Figure 1. Data source is an efficient hierarchical clustering algorithm, which can deal with large data clusters simultaneously. It is mainly used to divide the dataset to be processed into several small datasets to complete the quasi-clustering process of data.

**2.3. Core Ideas and Algorithms.** Clustering is the process of dividing a collection of physical or abstract objects into multiple classes composed of similar objects. The cluster generated by clustering is a collection of data objects that are similar to each other in the same cluster and different from the objects in other clusters. As the saying goes, "things come together in groups, people come together in clusters," and classification problems can be found everywhere in the social and natural sciences. Cluster analysis, also known as cluster analysis, is a statistical analysis method to study the problem of classifying samples or indicators.

Although clustering and classification have some commonality in that they are both aimed at classifying data from a given dataset, the difference is that clustering has a learning process, while classification simply divides the existing dataset into different categories according to the specified needs.

The number of categories for clustering can be specified or obtained automatically based on algorithms, and it is up to the analyst to decide which approach to use based on dif-

ferent business needs. The current development of clustering technology is promising and covers a wide range of fields, including statistics, data mining, machine learning, and marketing. Clustering analysis, as an important branch of data mining, is a popular research topic.

K-Means clustering is a classical bottom-up clustering method. It is characterized by a very fast clustering speed compared to other clustering algorithms even with a very large amount of data and a relatively simple execution process. However, the disadvantage is that the K-value must be specified before clustering, and the analyst usually does not know how many classes should be clustered in the actual application. This requires us to select different K-values for multiple clustering, which is solved by selecting the best clusters.

K-Means clustering is used to divide the existing dataset with  $n$  samples into clusters with high similarity. The clustering process can be roughly described as follows:  $k$  samples are randomly selected as clustering centers, the remaining ones calculate the distance from each sample to each center, then the object is assigned to the center closest to it, the center of each new cluster is recalculated, and the above process is repeated until the convergence condition is satisfied. The usual convergence condition is that the centroids no longer change or a certain number of iterations are reached. The squared error criterion is generally used, and the relevant definition is as follows:

$$E(n) = \sum_{i=1}^k \sum_{P \in Q} (\lambda t)^{n+1} e^{\lambda t}. \quad (1)$$

$E$  is the sum of the squared errors of all objects in the database,  $P$  is the point in space, and  $\bar{m}$  is the average of cluster  $i$ . The objective function is to make the generated clusters as compact and independent as possible, and the distance measure is the Euclidean distance, but it is possible to use other distance measures if desired.

$$P = e^{-2\pi T^2} = e^{-2G}. \quad (2)$$

The second-order clustering algorithm is a common hierarchical clustering algorithm, mainly used in the two major fields of cross-sectional and data mining of multivariate statistics. Compared with the K-Means clustering algorithm, the biggest advantage of the two-step clustering algorithm is that it can determine the K-value of clusters by itself, and there is no need to determine the best clustering class by different K-values, which eliminates many steps of manual discrimination.

But the disadvantage lies in the fact that because the categories are determined by the algorithm itself, without combining the actual needs, the clustering results often fail to achieve the expected results. The choice of the two algorithms needs to be made by analysts according to different needs and application scenarios. The following is a brief description of the algorithmic process of the two-step clustering algorithm for your reference.

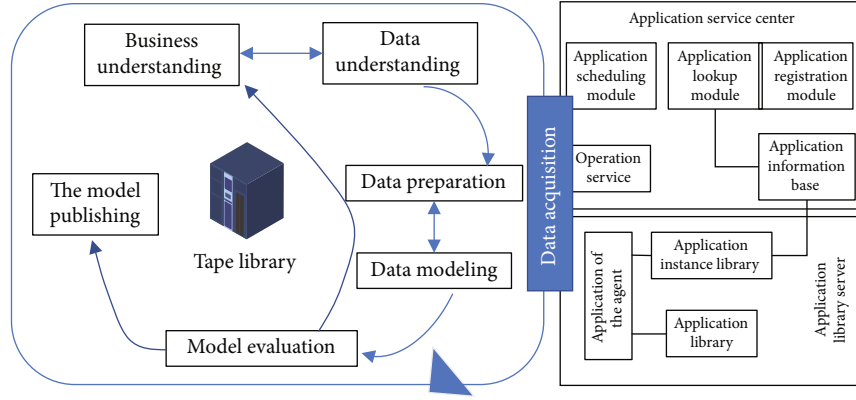


FIGURE 1: The 6 stages of CRISP.DM.

Input: the dataset with  $n$  samples and the number of clusters  $k$ .

Output:  $k$  clusters and minimizes the squared error criterion.

Steps.

- (1) Arbitrarily select  $k$  samples as initial cluster centers
- (2) Repeat the above steps
- (3) Calculate the mean of the samples in the cluster and reassign the values of the samples to the most similar cluster
- (4) Updating the cluster means, i.e., calculating the mean of the samples in each cluster
- (5) Until no more changes occur

ALGORITHM 1:

*Step 1.* Quasi-clustering process. The main work of this step is to apply a comprehensive hierarchical clustering algorithm BIRCH (Balance Iterative Reducing and Clustering using Hierarchies), which is an efficient hierarchical clustering algorithm capable of handling large-scale data clusters at one time. It is mainly used to divide the dataset to be processed into many small datasets and complete the quasi-clustering process of the data.

*Step 2.* Secondary clustering process. This step uses the log-likelihood function to calculate the distance between samples and implements secondary clustering based on the small dataset preprocessed in the previous step. The recursive algorithm merges the input subclusters using the “hierarchical coalescence method” until the last cluster includes all the datasets. The second-order clustering algorithm is integrated in Clementine, and the user only needs to select the appropriate model node to use. The algorithm uses probability-based distances as a measure function:

$$d(k, n) = (\lambda t)^{n+1} e^{\lambda t} + \xi_i, \quad (3)$$

$$\zeta_n = H_{N,v} \left( \sum_{i=1}^k \sum_{P \in Q} (\lambda t)^{n+1} e^{\lambda t} + Q_n \right), \quad (4)$$

$$\tilde{E}(n) = \sum_{i=1}^{I_k} \sum_{P \in Q} (\lambda t)^{n+1} e^{\lambda t} \log_n \frac{N_v}{N_n}, \quad (5)$$

where KA is the number of input domain ranges, KB is

the number of input domain symbols, KL is the number of categories of the  $k$ th symbolic domain of the input, VN is the number of records in cluster  $V$ , VKLN is the number of records belonging to the  $k$ th symbolic domain of the  $l$ th category in cluster  $V$ ,  $2k$  is the estimated deviation of the  $k$ th continuous variable from all records, and  $2vk$  is the estimated deviation of the  $k$ th continuous variable from the  $V$ th cluster.

RFM model is a means for enterprises to achieve database precision marketing, by setting three core indicators to carry out customer segmentation. In most cases, management decision-makers tend to focus only on the single dimension of sales amount and believe that those with high spending amount are important value customers and should be given key marketing. However, there are many customers with large single purchase amounts, but the frequency of sales is very low, and some even come into the store once in a few years; these customers can basically be judged as lost customers. For different customer groups, we should use differentiated marketing strategies, if the marketing of lost customers and important value customers using the same marketing approach will result in a waste of corporate resources. Currently, this kind of rough marketing model is gradually improving; the United States after a lot of research data shows that the following three factors constitute the key factors can be customer segmentation. They are freshness (the latest consumption time), consumption frequency (the number of times of consumption over a period of time), and consumption amount (the total amount of consumption over a period of time).

### (1) Freshness (last consumption time)

Freshness refers to the most recent consumption time of the customer, which category of goods the customer has recently purchased, whether the category has changed compared to the previous purchase of goods, what is the reason for the recent change in consumption habits, if the change from the purchase of daily goods to frequent purchase of baby products can be determined that the customer has a baby at home recently, and the change to the purchase of home building materials can be determined that the customer recently purchased a new house has the demand for decoration. For customers in different stages of life, merchants need to dig out customers' recent needs from purchase data and push the latest information and promotion information of products that customers are concerned about according to their consumption needs. Send targeted information through data analysis, so that the customer's acceptance of the information will be greatly enhanced. Do not make customers feel that you are sending information only to sell products; you need to reduce the customer's aversion to promotional information.

Once the customer becomes disgusted with the merchant, he or she will not go through the merchant's catalog book, the cell phone sets up SMS blocking, and the contact channel between the merchant and the customer will be cut off. If you send relevant information after three months or even six months, you will also miss the best marketing time period, and customer acceptance will be reduced. The actual situation shows that the acceptance of customers with recent consumption time is inevitably higher than that of customers with long consumption time. Through data analysis to dig out the information behind the customer, form effective interaction with the customer, repeated contact, so that the customer always feel the merchant's concern for him, enhance the customer's goodwill towards the merchant, and enhance the customer experience. Then, the customer's arrival rate will certainly be improved, and with the merchant's customized marketing strategy, the transaction rate will be significantly increased. By increasing the frequency of contact with the customer, the customer's arrival rate will be improved, thus increasing the closing rate.

### (2) Consumption frequency

This is the actual number of times a customer has visited the store over a period of time. These customers can be considered separately in the actual analysis process. The characteristics of these customers are that they love to take advantage of small bargains and will not come to the store without the merchant's activities and will only choose cheap goods, which will not form related sales and upselling. However, most of the customers with high consumption frequency are customers who often buy with high satisfaction to the merchant. The merchant's brand influence and service quality are what keep customers spending at high frequencies and loyalty. Customers have a limited time to buy, and increasing the frequency of customers' consumption will, to

a certain extent, grab market share from competitors and increase their own market share.

### (3) Spending amount

Spending is usually the most important metric for management and decision-makers, as it directly affects revenue and profitability. According to the traditional "two to eight" rule, 20% of customers contribute 80% of the company's revenue. It can be seen that how to distinguish the important value customers is very important to enterprises; in the case of limited resources and marketing expenses, you can give priority to 20% of the important value customers for precision marketing, because these customers are the main contributors to corporate revenue, as shown in Figure 2.

When classifying customers, we need to consider the above three core indicators and set different weights for the three indicators according to the actual situation of our own enterprise when using the RFM model in Clementine, and the weights can be adjusted by referring to the experienced values of marketing personnel.

## 3. Experimental Design

Freshness (last consumption time), consumption frequency, and consumption amount are three very important indicators in customer value assessment and customer relationship management, which have a strong guiding meaning for our daily marketing activities, customer maintenance, and enhancing customer loyalty. The freshness is the most important of them. Using RFM model in Clementine, we can transform these three variables as long as the analyzed dataset has three indicators: unique customer identification, last time, and spending amount. Each of these variables is given a different weight according to the actual situation, and then, the customer value is evaluated based on the RFM score. We classify customers into  $5 \times 5 \times 5 = 125$  categories, as shown in Figure 3, and analyze their data to customize marketing strategies.

Referring to the amazon-meta dataset of Amazon, the amount of review data shown in Dataset statistics is 7781990, which is the total sum of review statistics table of products, including 42919 undownloaded data, 145827 invalid data, the total number of reviews (count), the average value of evaluation scores (AVG\_rating), the average value of the number of votes (AVG\_votes), the average value of the number of customer reviews (AVG\_votes), and the average value of the number of customer reviews (AVG\_votes). For (AVG\_helpful), the total number of data is 1555171.

Companies must carry out marketing activities before the consumer habits and buying behavior data mining, to understand the customer in order to develop marketing strategies. Traditional market research methods have been tested in practice and have great limitations. The customer base is narrow and unrepresentative. Data mining makes up for this shortcoming. Big data mining based on social network evaluation data can be carried out from the following aspects.

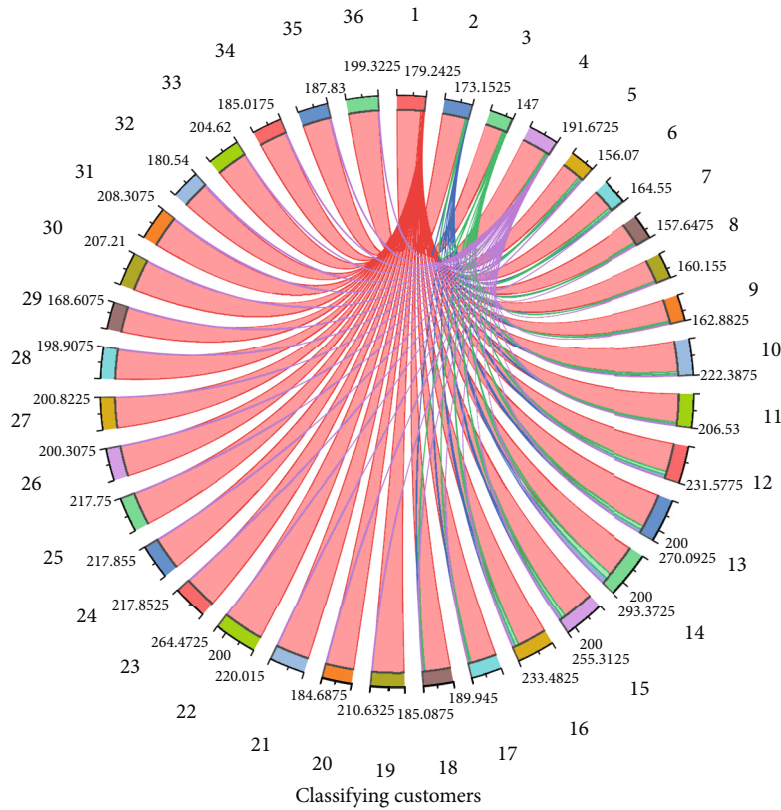


FIGURE 2: RFM model.

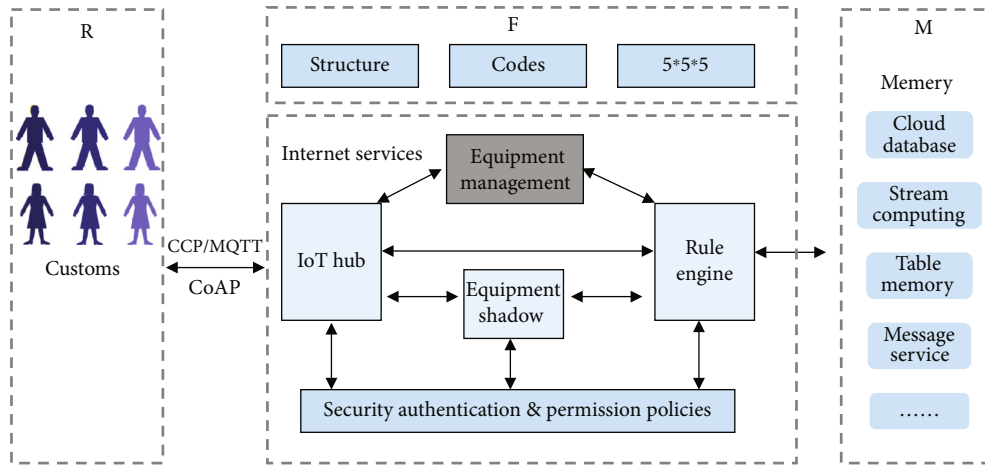


FIGURE 3: RFM coding structure.

- (1) Preprocessing of data and clustering using different clustering algorithms
- (2) Analysis of the differences between the customer groups formed by the clustering results
- (3) From the differentiation of data, find out the characteristics of customer reviews and the interesting behavior of customers in purchasing goods and conduct a series of analysis and summary, so as to provide users and enterprises with more targeted and valuable information

## 4. Results and Analysis

4.1. Quality Node Mining. RFM model: R (recency) indicates the most recent purchase time of the customer, F (frequency) indicates the frequency of the customer’s consumption in a period of time, and M (monetary) indicates the amount of the customer’s consumption in a period of time. The original fields analyzed are three: customer ID (unique identification of customers), consumption time (date format), and consumption amount, which are processed by data mining software to obtain RFM scores by weighting the three indicators,

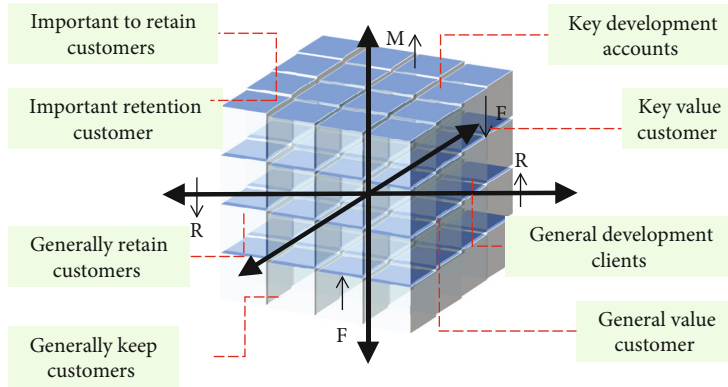


FIGURE 4: Traditional RFM model and Amazon’s RFM model.

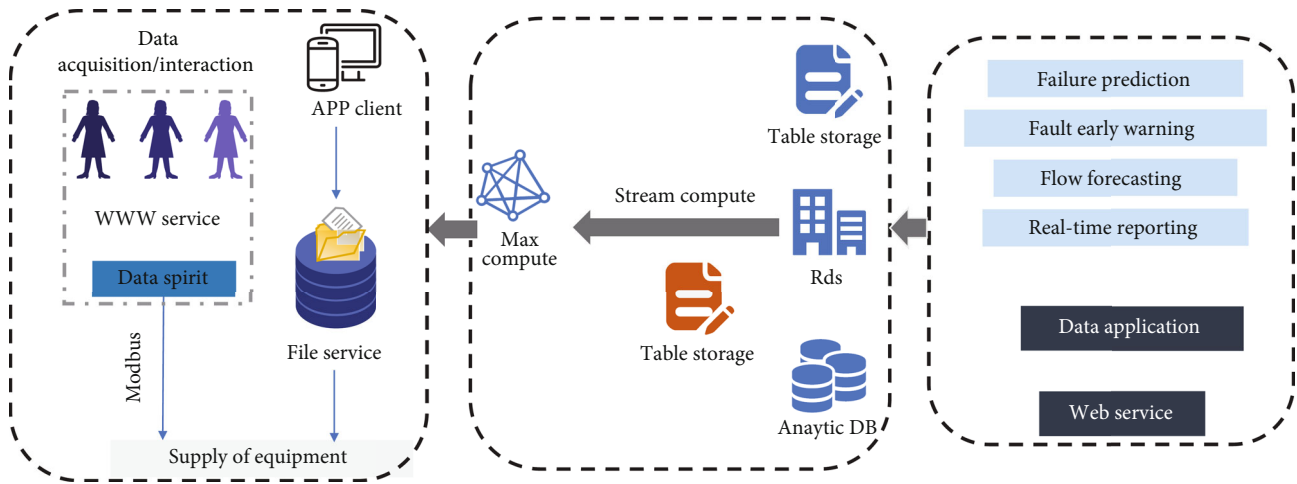


FIGURE 5: Customer segmentation based on RFM model.

completing customer class classification, and sorting the obtained RFM scores to implement accurate database marketing for different customer classes. Figure 4 shows the transformation of traditional RFM model into Amazon’s RFM model.

RFM model is a data processing method, and in this paper, we found that the traditional RFM model classifies to form 125 customer categories, which is too segmented customers and difficult to operate in practical application. We need to adopt a new clustering method for customer segmentation based on RFM model, and the characteristics and behavioral habits of segmented customer groups need to reach our expected clustering under over.

We continue to use Clementine to cluster the three fields of R, F, and M. The clustering analysis mainly uses the following: K-Means and two-step algorithms. Before clustering, we found that directly using the three variables of R, F, and M for clustering, the measurement scales between the variables are very different and cannot achieve the expected effect, so we need to have descaling. In addition, the weights of R, F, and M should be different because of the differences in the importance of these three indicators in the displayed evaluation system. By comparing the functionality of a topic, the following statistics can

be obtained for the number of times a given user has posted each topic. Here, we do not use a weighting method for the three variables (in practice, an expert or a corresponding marketer is needed to evaluate them), and through continuous testing and evaluation, we can choose the specific clustering method and the number of clusters, and also, we need to compare which of the two algorithms is more desirable. The RFM model-based customer segmentation quadrant is shown in Figure 5.

Drawing a tree graph starts by constructing a specific function whose return value is the overall height of the given cluster. It is important to know the overall height of the clusters when determining the overall height of the graph and placing the different node positions. If the cluster is a leaf node, its height is 1; otherwise, the height is the sum of all branch heights. In addition, the overall error of the root node must be known. That is because the length of the lines is adjusted accordingly to the error of each node, so a scaling factor is generated based on the total error value. The depth of error of a node is equal to the maximum possible error of each branch to which it belongs.

The drawdendrogram function creates an image with a fixed width and a height of 20 pixels for each of the final generated clusters. The scaling factor is obtained by

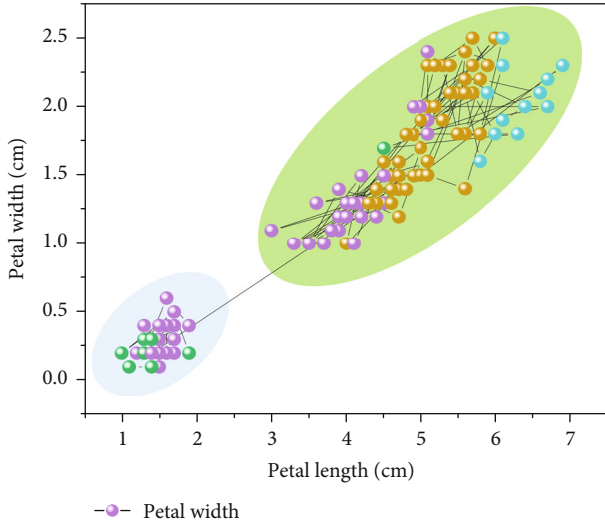


FIGURE 6: Clustering similarity.

dividing the fixed width by the total depth value. This function creates the corresponding draw object for the image and then calls the drawnode function at the position of the root node and places the node at the left center of the entire image. When the drawnode function accepts a cluster and its position as input parameters, the function takes the height of the child nodes and calculates their exact position in the picture and then connects them with lines. The diversity of Web data forms has brought new challenges to data mining. Combining clustering, support vector machine, neural network, and other database technologies, in the massive graphical and complex spatiotemporal data, is the main effort and development direction of future data mining. There are two horizontal lines and one longer vertical line. The length of the horizontal line is determined by the error condition in the clustering. The length of the line is related to the clustering result, the more different the two clusters are combined, the longer the line is drawn, and if the two clusters are more similar, the shorter the line is drawn, as shown in Figure 6.

According to the result of the above figure, there are eight levels. First, topic 4 and topic 5 are clustered, assumed to be cluster 1, and topic “Little Times 3” and topic “120 years of Wudao” are clustered, assumed to be cluster 2. At this point, topic 1 and topic 2 are clustered to form cluster 3 and so on, and finally, all the clusters are merged into one class. Finally, all clusters are combined into one class. And from Figure 7, we can see that after topic 3 and other topics are clustered into one class, the topic “Where did flowers go” is much farther away from the aggregation point, so we can conclude that the classes formed by other topics are closer to the aggregated class. There are 28 topics in the above figure, and 31 topics are extracted, so we can conclude that the remaining 3 topics are outliers and cannot be clustered with other classes.

The function will return a value between -1 and 1, and the number of evaluations must be greater than or equal to 0, so the final value returned is a value between 0 and 1. A value of

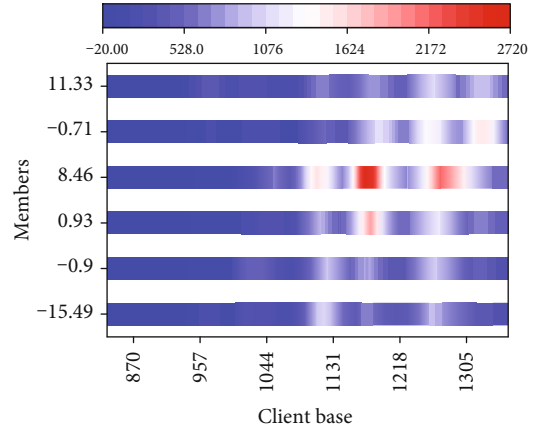


FIGURE 7: Clustering algorithm viewer.

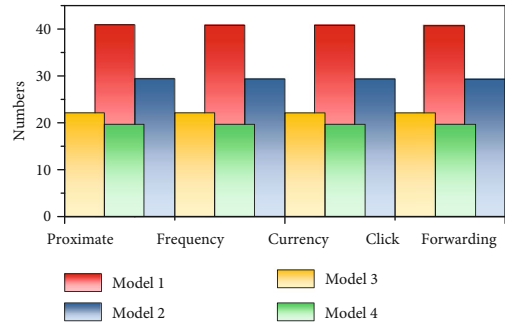


FIGURE 8: Differences in metrics between categories.

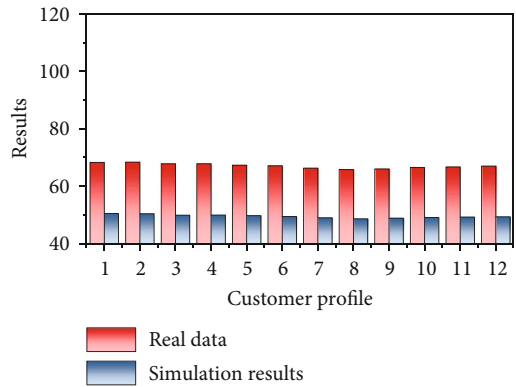


FIGURE 9: Distribution of the number of customers by category.

1 means that each user has the same number of posts for both topics. With the function of comparing topics, the following statistics can be made on the number of times a given user has posted on each topic, and if the number of times is higher, it means that the user has a higher interest in the topic, and the interest can be expressed by scoring the evaluation, as shown in Figure 8. However, there may be a user who is particularly enthusiastic about a topic, which will affect the final evaluation result. In order to solve this problem, the authors propose a weighted evaluation value to score the topic, and then, the ranking of the topic evaluation will be obtained. Finally, the similarity obtained is multiplied by the



evaluation value obtained for each topic to reach the final result. The program obtains an ordered list of topics that are of similar interest to a given person and that have not been tweeted by that user. The specified person is randomly generated in the program.

Since the 31 popular topics and the users involved in each topic were extracted in the previous section, it is possible to recommend topics with high similarity to users based on this data, which is also known as collaborative filtering technique. Collaborative filtering algorithms usually delineate a large group of people and then search for them to find a group of people with similar tastes. The algorithm focuses on the content that the group likes and ranks the content together to get a list of recommendations, as shown in Figure 9. After collecting the data, the number of tweets posted by each user on each topic is determined, and then, the similarity between topics is determined based on the number of tweets posted by each user on each topic, which can be used to calculate the similarity between each topic and other topics.

## 5. Conclusion

Discovering the hidden value in the data is the main purpose of data mining applications. In recent years, the rise of e-commerce networks, the development of people's online shopping habits, the development of social networks is remarkable, and people's interactive behavior on the Internet has formed a large amount of social network data, the traditional statistical analysis methods have limitations for dealing with big data, and the hidden patterns and values in the data can be discovered with the method of big data mining which is a new application trend. In this paper, we combine traditional statistical analysis methods and data mining algorithms to analyze Amazon's customer evaluation data. Around this work, this paper finds that the customer evaluation data of Amazon is analyzed by cluster analysis method, and two different clustering algorithms are applied to classify customer groups using four evaluation indexes of customers, and the results of clustering are compared and evaluated, and the optimal clustering results are selected to make a customer portrait for Amazon's evaluation customers. In addition, this paper applied the clustering algorithm based on RFM model to classify the customers of Amazon customer evaluation data, observe the proportion of important value customers and whether there are missing growth customers, and provide relevant suggestions for customer development and maintenance.

## Data Availability

Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

## Consent

Informed consent was obtained from all individual participants included in the study references.

## Conflicts of Interest

We declare that there is no conflict of interest.

## References

- [1] D. Zhang, J. Yin, X. Zhu, and C. Zhang, "Network representation learning: a survey," *IEEE Transactions on Big Data*, vol. 6, no. 1, pp. 3–28, 2020.
- [2] Z. He, Z. Cai, and J. Yu, "Latent-data privacy preserving with customized data utility for social network data," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 1, pp. 665–673, 2018.
- [3] E. Kross, P. Verduyn, M. Boyer et al., "Does counting emotion words on online social networks provide a window into people's subjective experience of emotion? A case study on Facebook," *Emotion*, vol. 19, no. 1, pp. 97–107, 2019.
- [4] G. Xu, Y. Meng, X. Qiu, Z. Yu, and X. Wu, "Sentiment analysis of comment texts based on BiLSTM," *IEEE Access*, vol. 7, pp. 51522–51532, 2019.
- [5] M. Moessner, J. Feldhege, M. Wolf, and S. Bauer, "Analyzing big data in social media: text and network analyses of an eating disorder forum," *International Journal of Eating Disorders*, vol. 51, no. 7, pp. 656–667, 2018.
- [6] M. Balaanand, N. Karthikeyan, and S. Karthik, "Designing a framework for communal software: based on the assessment using relation modelling," *International Journal of Parallel Programming*, vol. 48, no. 2, pp. 329–343, 2020.
- [7] F. Ali, S. El-Sappagh, S. R. Islam et al., "An intelligent health-care monitoring framework using wearable sensors and social networking data," *Future Generation Computer Systems*, vol. 114, pp. 23–43, 2021.
- [8] J. R. Ragini, P. R. Anand, and V. Bhaskar, "Big data analytics for disaster response and recovery through sentiment analysis," *International Journal of Information Management*, vol. 42, pp. 13–24, 2018.
- [9] J. Kim and M. Hastak, "Social network analysis," *International Journal of Information Management*, vol. 38, no. 1, pp. 86–96, 2018.
- [10] M. Klöwer, D. Hopkins, M. Allen, and J. Higham, "An analysis of ways to decarbonize conference travel after COVID-19," *Nature*, vol. 583, no. 7816, pp. 356–359, 2020.
- [11] L. Liao, X. He, H. Zhang, and T.-S. Chua, "Attributed social network embedding," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 12, pp. 2257–2270, 2018.
- [12] J. A. Obar and A. Oeldorf-Hirsch, "The biggest lie on the internet: ignoring the privacy policies and terms of service policies of social networking services," *Information, Communication & Society*, vol. 23, no. 1, pp. 128–147, 2020.
- [13] K. K. Kapoor, K. Tamilmani, N. P. Rana, P. Patil, Y. K. Dwivedi, and S. Nerur, "Advances in social media research: past, present and future," *Information Systems Frontiers*, vol. 20, no. 3, pp. 531–558, 2018.
- [14] Y. Li, J. Fan, Y. Wang, and K.-L. Tan, "Influence maximization on social graphs: a survey," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 10, pp. 1852–1872, 2018.
- [15] L. Bode and E. K. Vraga, "See something, say something: correction of global health misinformation on social media," *Health Communication*, vol. 33, no. 9, pp. 1131–1140, 2018.
- [16] M. E. J. Newman, "Network structure from rich but noisy data," *Nature Physics*, vol. 14, no. 6, pp. 542–545, 2018.

- [17] Y. Dong, Q. Zha, H. Zhang et al., "Consensus reaching in social network group decision making: research paradigms and challenges," *Knowledge Based Systems*, vol. 162, pp. 3–13, 2018.
- [18] X. Zheng, J. Han, and A. Sun, "A survey of location prediction on Twitter," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 9, pp. 1652–1671, 2018.
- [19] C. Brell, C. Dustmann, and I. Preston, "The labor market integration of refugee migrants in high-income countries," *Journal of Economic Perspectives*, vol. 34, no. 1, pp. 94–121, 2020.
- [20] D. Holtz, M. Zhao, S. G. Benzell et al., "Interdependence and the cost of uncoordinated responses to COVID-19," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 117, no. 33, pp. 19837–19843, 2020.
- [21] D. Wang, B. K. Szymanski, T. Abdelzaher, H. Ji, and L. Kaplan, "The age of social sensing," *IEEE Computer*, vol. 52, no. 1, pp. 36–45, 2019.

## Research Article

# Precision Measurement for Industry 4.0 Standards towards Solid Waste Classification through Enhanced Imaging Sensors and Deep Learning Model

Leow Wei Qin,<sup>1</sup> Muneer Ahmad <sup>1</sup>, Ihsan Ali <sup>2</sup>, Rafia Mumtaz <sup>3</sup>,  
Syed Mohammad Hassan Zaidi <sup>3</sup>, Sultan S. Alshamrani <sup>4</sup>, Muhammad Ahsan Raza <sup>5</sup>,  
and Muhammad Tahir<sup>6</sup>

<sup>1</sup>Department of Information Systems, Faculty of Computer Science & Information Technology, Universiti Malaya, 50603 Kuala Lumpur, Malaysia

<sup>2</sup>Department of Computer System and Technology, Faculty of Computer Science & Information Technology, Universiti Malaya, 50603 Kuala Lumpur, Malaysia

<sup>3</sup>National University of Sciences and Technology (NUST), School of Electrical Engineering and Computer Science (SECS), Islamabad, Pakistan

<sup>4</sup>Department of Information Technology, College of Computer and Information Technology, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia

<sup>5</sup>Department of Information Technology, Bahauddin Zakariya University, Multan 60000, Pakistan

<sup>6</sup>Department of Computer Science, Abdul Wali Khan University, Mardan, 23200, KPK, Pakistan

Correspondence should be addressed to Ihsan Ali; [ihsanalichd@siswa.um.edu.my](mailto:ihsanalichd@siswa.um.edu.my)

Received 29 March 2021; Revised 20 April 2021; Accepted 27 April 2021; Published 17 May 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Leow Wei Qin et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Achievement of precision measurement is highly desired in a current industrial revolution where a significant increase in living standards increased municipal solid waste. The current industry 4.0 standards require accurate and efficient edge computing sensors towards solid waste classification. Thus, if waste is not managed properly, it would bring about an adverse impact on health, the economy, and the global environment. All stakeholders need to realize their roles and responsibilities for solid waste generation and recycling. To ensure recycling can be successful, the waste should be correctly and efficiently separated. The performance of edge computing devices is directly proportional to computational complexity in the context of nonorganic waste classification. Existing research on waste classification was done using CNN architecture, e.g., AlexNet, which contains about 62,378,344 parameters, and over 729 million floating operations (FLOPs) are required to classify a single image. As a result, it is too heavy and not suitable for computing applications that require inexpensive computational complexities. This research proposes an enhanced lightweight deep learning model for solid waste classification developed using MobileNetV2, efficient for lightweight applications including edge computing devices and other mobile applications. The proposed model outperforms the existing similar models achieving an accuracy of 82.48% and 83.46% with Softmax and support vector machine (SVM) classifiers, respectively. Although MobileNetV2 may provide a lower accuracy if compared to CNN architecture which is larger and heavier, the accuracy is still comparable, and it is more practical for edge computing devices and mobile applications.

## 1. Introduction

Industry 4.0 standards require cutting-edge solutions to prevent, reduce, and even eradicate solid waste to ensure a pollution-free and sustainable environment [1]. As defined

by the Environmental Protection Agency (EPA), municipal solid waste (MSW) is a trash from sources that include residential, commercial, and institutional locations, such as businesses, schools, and hospitals [2]. However, the definition of MSW by EPA does not include industrial, hazardous,

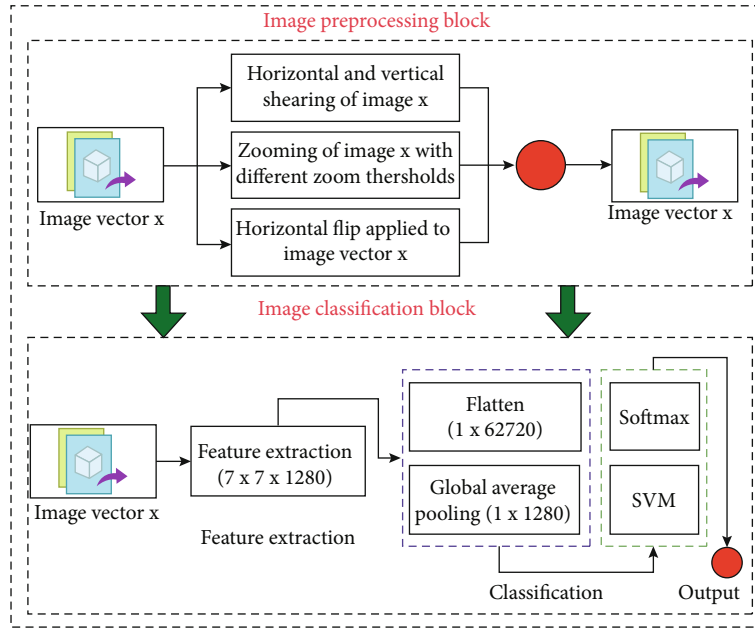


FIGURE 1: General architecture of the proposed classification model.

or construction and demolition (C&D) waste. MSW is locally handled by each municipality [3].

Municipal solid waste (MSW) can generally be classified into two categories: organic and inorganic categories [4]. To break it down further, organic waste consists of food waste and yard waste, while inorganic waste contains plastics, metal, cans, paper, glass, and others [5]. The waste composition differs from country to country as they are subjected to the influence of many factors, such as level of economic development, cultural norms, geographical location, energy sources, and climate. Inorganic waste such as plastics, paper, and aluminium increases while the relative organic waste decreases when a country urbanizes and populations become wealthier [6]. On the other hand, for low- and middle-income countries, organic waste would form the majority percentage in the urban waste stream, ranging from 40 to 85% of the total [7]. Paper, plastic, glass, and metal fractions increase in the waste stream of the middle- and high-income countries [8].

Significant advancement in the standard of living will inevitably increase MSW generation [9]. This is because when the population's spending power increases, consumption of goods and services increases as well [10]. When consumption increases, the population will demand more resources like water, energy, minerals, and land which will give rise to the amount of waste generated [11]. Along with a higher standard of living, the increasing population from rural areas to urban areas had also contributed to a higher MSW generation rate [12]. Based on an estimation by World Bank, currently, 1.3 billion tons of waste is generated annually over the world, and this amount will increase to 2.2 billion tons annually by 2025 [13]. Thus, if waste is not managed properly, it would bring about an adverse impact on health, the economy, and the global environment. It was

```

Loop (1): a in range (until the width of vector x)
Loop (2): b in range (until the height of vector x)
a' = width (vector x) - a - 1
End Loop (2):
End Loop (1):

```

CODE 1

also reported that improper management of waste will result in a cost higher than what it would have cost if waste was managed appropriately in the first place. As such, like it or not, all parties will need to assume more responsibility for waste generation and disposal, specifically, product design and waste separation [14].

Considering that waste is mainly a by-product of consumer-based lifestyles that is the main driver of the world's economies, in most cases, the fastest way to reduce waste generation is to reduce economic activity [15]. However, a reduction in economic activity would not be attractive [16]. With that, EPA has recognized MSW recycling as the second most environmentally sound strategy for dealing with urban waste [17].

The recent ICT-driven solutions have played a substantial role in MSW management. The state-of-the-art technologies, i.e., computer vision, edge computing, IoT, machine learning, and deep learning, have been proven to be an enormous support to achieve industry 4.0 standards [18]. The current industry employs the latest MSW management tools that are economically and computationally inexpensive and efficient [19].

The accurate classification of MSW employing machine learning models has considerably benefited the MSW

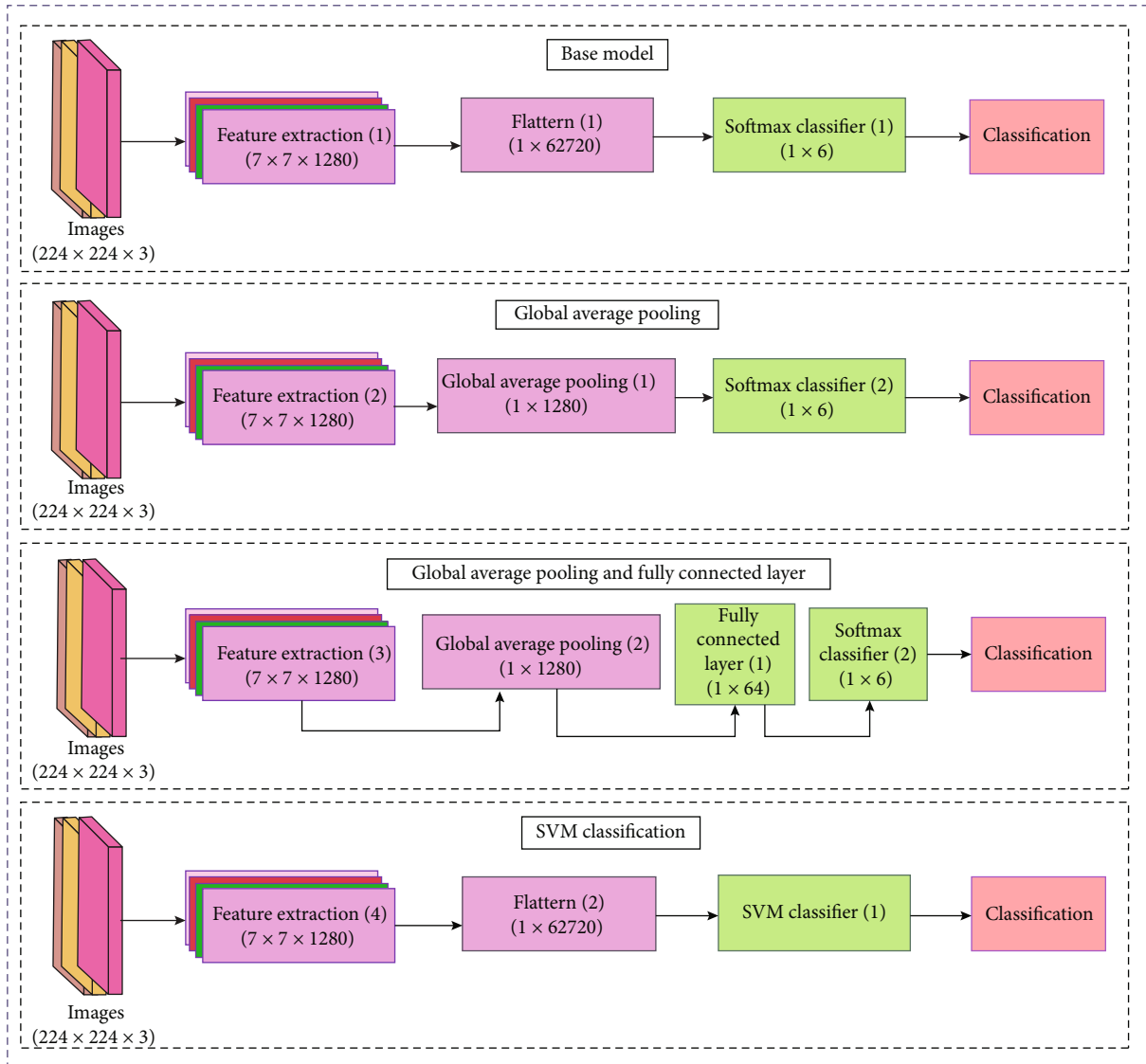


FIGURE 2: Component-based scenarios of the proposed classification model.

stakeholders. [20] proposed machine learning as a useful tool for accurate prediction and management of solid waste in Canada based on socio-economic and demographic variables. The study applied decision trees and neural networks to build the models. Ultimately, the study developed an integrated framework for accurate classification and management of municipal solid waste of 220 municipalities of Ontario. Similarly, [21] developed algorithms to understand the patterns of solid waste generations using machine learning and small area estimation techniques. The study incorporated the prediction of MSW to quantify the future estimates of waste generation to ease the recycling of waste materials for reusability. With these advancements, the ML algorithms, i.e., regression classification, naïve Bayes classification, support vector machine classification, decision trees, KNN, random forests, and CNN, have bestowed computationally inexpensive and fruitful solutions for MSW classifications. As the main advantages of these solutions, efficient resource utilisation has given economical solutions to local

and urban waste management [22]. Thus, significantly assisted in achieving a sustainable and pollution-free environment and enhanced the quality and standards of lives of common people [23].

## 2. Related Works

The study observes many citations related to forecasting future solid waste generation and its quantification, e.g., [24]. Similarly, another good number of citations can be observed for MSW using sensors and IoT devices, e.g., [25, 26]. On contrary, the study could find a little citable work related to optimization or enhancement of solid waste classification methods developed for edge computing devices or other lightweight applications. Among the recent citations, [27] developed an automated recognition system employing a deep learning algorithm for the classification of solid waste objects as biodegradable and nonbiodegradable.

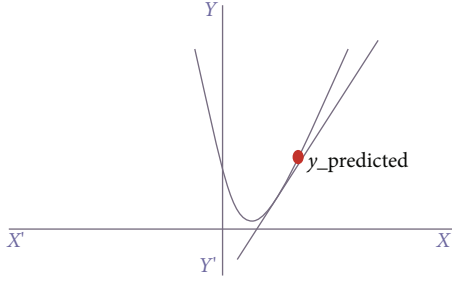


FIGURE 3: Gradient descent in the proposed approach.

```

Loop var i in range(10000):
y_predicted = m_curr * x + b_curr
md = -(2/n) * sum(x * (y - y_predicted))
yd = -(2/n) * sum(y - y_predicted)
m_curr = m_curr - rate * md
b_curr = b_curr - rate * yd
End Loop

```

CODE 2

Moving on, [28] used two models, namely, support vector machine (SVM) and CNN for waste image classification. SVM was used based on the principle of processing features in space higher than the original features space. Using a higher feature space, the data was divided into different categories using a hyperplane. SVM computed the selection of planes to develop the largest separation of different categories among the data. For CNN, the study had implemented an architecture that was similar to AlexNet but smaller due to computational constraints. At that point, there were no publicly available datasets, and the study initially found Flickr Material Database and Google Images. However, the images could not accurately represent the state of recycled goods. Such issues resulted in the creation of a new dataset named TrashNet that contained around 2400 images with six different classes. Data augmentation techniques such as random rotation, random brightness control, random translation, random scaling, and random shearing were applied to increase the dataset size. Surprisingly, SVM performed better than CNN even though SVM is a much simpler model. The authors believed that the poor performance of CNN was due to suboptimal hyperparameters.

Another work done by [29] proposed an image classification model related to MSW management. The study employed a camera to picture the solid waste and categorize it according to five defined waste categories trained on a dataset of 400-500 images. In a similar context, [30] employed SVM for the supervised classification of solid waste. The study segregated the solid waste images into recyclable or nonrecyclable.

The study done by [31] leveraged the transfer learning using a pretrained network utilizing AlexNet, GoogLeNet, VGG-16, and ResNet to classify waste images with different categories with the TrashNet dataset. In the final output layer

TABLE 1: Training result (10-fold crossvalidation) for MobileNetV2 using Softmax classifier.

Fold	Accuracy	Loss	Precision	Recall
Fold 1	91.832	1.605	0.915	0.914
Fold 2	92.574	1.112	0.911	0.910
Fold 3	87.129	1.704	0.912	0.910
Fold 4	85.396	3.124	0.915	0.913
Fold 5	91.584	1.323	0.917	0.916
Fold 6	91.337	1.506	0.915	0.914
Fold 7	89.851	1.831	0.915	0.914
Fold 8	84.901	2.714	0.913	0.912
Fold 9	85.856	2.013	0.906	0.904
Fold 10	88.089	2.052	0.923	0.922
Average	<b>88.855</b>	<b>1.898</b>	<b>0.914</b>	<b>0.913</b>

of the CNN, different classifiers were used, namely, Softmax and support vector machine (SVM). The study segregated training and testing data into equal halves without performing any further data augmentation.

[32] used VGG-19 for the waste image classification. The study used data augmentation such as shear, rotation, zooms, and shifts in addition to resizing images to  $224 \times 224 \times 3$  dimensions before feeding the network. The classification layer of pretrained VGG-19 was removed and replaced by a fully connected layer of 256 neurons with ReLU activation function with a dropout of 0.6 and batch normalization. The model was trained using a batch size of 32, and the authors observed a learning rate between 0.001 and 0.0001 bestowed better results. The lower validation accuracy depicted the overfitting of the model even though [31] and [32] had managed to train the CNN model with high accuracy. The CNN architecture used, e.g., AlexNet, contained about 62,378,344 parameters, and over 729 million FLOPs required classifying a single image. A deep architecture as AlexNet or VGG increases the model complexity. As a result, the deployment of these huge CNN models is often unaffordable for common computers and mobile devices. Thus, this had given rise to the need for a lightweight CNN architecture that could still achieve good results when compared to huge CNN models such as AlexNet and VGG.

To create a lightweight deep neural network that could be used for embedded vision applications and mobile applications, [33] proposed a new architecture named MobileNet which is based on depthwise separable convolution. Depthwise separable convolution is a depthwise convolution followed by a pointwise convolution. Using a depthwise separable layer, the number of multiplication operations required is reduced significantly compared to normal convolution. Although being smaller in size, MobileNet has managed to achieve good results compared to other popular models such as VGG and GoogLeNet.

Following the success of MobileNet, [34] proposed an improved version of MobileNet named MobileNetV2. The study incorporated two changes to MobileNet. The first change is to incorporate an expansion layer before the depthwise convolution. This module takes a low-dimensional

Index	Cardboard	Glass	Metal	Paper	Plastic	Trash
Cardboard	67	0	1	12	0	1
Glass	0	77	7	3	13	1
Metal	0	13	64	1	3	1
Paper	2	0	2	107	1	7
Plastic	0	18	2	5	69	3
Trash	1	0	1	4	2	20

FIGURE 4: Confusion matrix (test set) using Softmax classifier.

TABLE 2: Precision and recall (test set) for MobileNetV2 using Softmax classifier.

Category	Precision	Recall
Cardboard	0.957	0.827
Glass	0.724	0.752
Metal	0.797	0.768
Paper	0.826	0.916
Plastic	0.769	0.722
Trash	0.613	0.679
Average	0.781	0.777

compressed representation which is first expanded to high dimensions and later on pass-through  $1 \times 1$  convolution to project it back to a low-dimensional representation. Hence, the expansion layer will always have more output channels than the input. The second change proposed by [34] is to include a residual block similar to ResNet. Each layer has batch normalization, and the activation function used is ReLU6. Only the output of the projection layer does not have an activation function. This is because the projection layer produces a low-dimensional output, and applying nonlinearity on it will destroy useful information.

The magnitude of expansion is defined by the expansion factor, which is known as the width multiplier,  $\alpha$  in MobileNet. For example, if there is an input with 24 channels and  $\alpha$  is 6, the expansion layer will create a new feature with  $24 \times 6 = 144$  channels. Subsequently, the depthwise convolution will be applied followed by the projection layer ( $1 \times 1$  convolution) which will project the 144 channels back into a smaller number, e.g., 24. The expansion and projection module is named as bottleneck layer [35]. Ideally, the number of channels should be large so that more information can be extracted. Combining both the expansion layer and projecting layer enables the model to extract more information while keeping the feature dimension rather small.

### 3. Methodology

In this research, we intend to develop an enhanced image classification model based on the lightweight feature extraction model MobileNetV2 to classify waste images into differ-

TABLE 3: Training result (10-fold crossvalidation) for MobileNetV2 using SVM classifier.

Fold	Accuracy	Precision	Recall
Fold 1	94.307	0.954	0.943
Fold 2	93.812	0.942	0.923
Fold 3	94.802	0.956	0.937
Fold 4	91.584	0.926	0.901
Fold 5	94.307	0.931	0.939
Fold 6	95.050	0.959	0.935
Fold 7	92.574	0.934	0.907
Fold 8	95.050	0.948	0.954
Fold 9	95.782	0.962	0.944
Fold 10	95.533	0.960	0.927
Average	<b>94.280</b>	<b>0.947</b>	<b>0.931</b>

ent categories according to their material. The prevailing citations of referred work in waste image classification employed convolution neural network models that were found to be larger and computationally expensive.

Let us consider a collection  $S = \{x_1, x_2, x_3, x_4, \dots, x_n\}$ ,  $n \leq N$  contains  $n$  images from  $N$  image space. Here, an image  $\mathbf{x}$  is a vector such that  $\mathbf{x} \in R^D$  with  $D$  number of dimensions. Let us consider  $\mathbf{x} \in R^{A \times B \times C}$  where each  $\mathbf{A}$ ,  $\mathbf{B}$ , and  $\mathbf{C}$  represent the row vector, column vector, and color vector having specific indices  $i, j, k$  defined by  $0 \leq i \leq D$ ,  $0 \leq j \leq D$ , and  $0 < k \leq 3$ .

Figure 1 depicts a general architecture of the proposed classification model. The input vector  $\mathbf{x}$  is passed through a series of preprocessing steps explained below. The feature extraction of output vector  $\mathbf{x}'$  assists in accurate classification.

We breakdown the preprocessing steps of image vector  $\mathbf{x}$  as follows:

- (1) We apply the horizontal and vertical shearing to image  $\mathbf{x}$  with coordinates  $a$  and  $b$  as described in Equations (1) and (2)

$$\begin{pmatrix} a' \\ b' \end{pmatrix} = \begin{pmatrix} a + mb \\ b \end{pmatrix} = \begin{pmatrix} 1 & m \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}, \quad (1)$$

Index	Cardboard	Glass	Metal	Paper	Plastic	Trash
Cardboard	69	0	3	9	0	0
Glass	0	77	9	2	13	0
Metal	0	7	68	3	4	0
Paper	3	1	1	108	1	5
Plastic	2	6	2	2	82	3
Trash	0	0	2	4	3	19

FIGURE 5: Confusion matrix (test set) using SVM classifier.

$$\begin{pmatrix} a' \\ b' \end{pmatrix} = \begin{pmatrix} a \\ mx + b \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ m & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}. \quad (2)$$

The horizontal and vertical shear displace the point of interest to adjust image  $\mathbf{x}$  for the desired shear.

- (2) We apply the zooming to image vector  $\mathbf{x}$  and achieve another image  $\mathbf{x}'$  such that  $\mathbf{x}$  is ( $z * 10\%$ ) of vector  $\mathbf{x}$ . Here,  $z$  is a particular point in image  $\mathbf{x}$  having coordinates  $a$  and  $b$ . We can assume that zooming will give us zoomed image with points  $(za, zb)$  that specify zooming points with a displacement  $z$ . It is important to note down that various zooming can be applied to predict the network better
- (3) The horizontal flip applied to image vector  $\mathbf{x}$  with coordinates  $a$  and  $b$  produces an image  $\mathbf{x}$  with coordinates  $a'$  and  $b'$  such that  $a' = \text{width}(\text{vector } \mathbf{x}) - a - 1$  while  $b' = b$ . We can describe this phenomenon as follows

Let us define the component-based scenario of our convolution neural network as

$$x^P \rightarrow w^P \rightarrow x^Q \rightarrow w^Q \rightarrow x^Z \rightarrow w^Z \rightarrow o^T. \quad (3)$$

Here, vector  $\mathbf{x}$  is an input image that is handled by the first layer of the network with weight  $\mathbf{w}$  (the vector that processes the input  $\mathbf{x}$  by applying weight  $\mathbf{w}$ , and the result of this computation is forwarded to the second layer). Next, the vector  $\mathbf{x}^Q$ , which was the output of the first layer, now serves as an input for the second layer managed by weight  $\mathbf{w}^Q$ . This process keeps on going for a defined number of layers until the outcome  $\mathbf{O}^T$  (where  $T$  represents the number of output channels, i.e., prediction of an image in two or more classifications) is reached. It is important to describe that  $\mathbf{o}^T$  is a trade-off between the last predicted value of vector  $\mathbf{x}^Z$  with the desired target (e.g.,  $M$ ) defined by Euclidean distance  $d = 1/2 |M - \mathbf{o}^T|^2$ . The distance  $d$  can also be measured using other measures, e.g., Manhattan distance.

In an ideal scenario, the number of channels should be larger so to extract more information. Integration of two layers, i.e., expansion and projection layer, can significantly enable the model for extracting additional information. This

TABLE 4: Precision and recall (test set) for MobileNetV2 using SVM classifier.

Category	Precision	Recall
Cardboard	0.932	0.840
Glass	0.831	0.733
Metal	0.791	0.878
Paper	0.844	0.908
Plastic	0.794	0.835
Trash	0.840	0.750
Average	0.839	0.824

TABLE 5: Training result (10-fold crossvalidation) global average pooling using Softmax classifier.

Fold	Accuracy	Loss	Precision	Recall
Fold 1	88.861	0.323	0.909	0.820
Fold 2	87.624	0.384	0.913	0.816
Fold 3	85.891	0.437	0.914	0.819
Fold 4	82.426	0.499	0.918	0.828
Fold 5	89.604	0.302	0.911	0.813
Fold 6	89.109	0.373	0.913	0.821
Fold 7	88.614	0.301	0.912	0.819
Fold 8	85.891	0.349	0.913	0.819
Fold 9	89.578	0.336	0.911	0.817
Fold 10	88.337	0.372	0.914	0.820
Average	<b>87.594</b>	<b>0.368</b>	<b>0.913</b>	<b>0.819</b>

holds an added value of keeping the feature dimension even small in size. The activation function ReLU6 can also be significant in this context.

Figure 2 showed the summarized components based on various scenarios of the proposed classification model. The adjustment of weights is essentially important to reduce the error; a gradient descent can be adopted as described in Figure 3 and below algorithm.

Let us consider slope and intercept to be defined by  $m_{\text{curr}}$  and  $b_{\text{curr}}$ , respectively, and both initialized to 0. Let us assume the initial learning rate as 0.01. Let  $n$  represent the length of vector  $\mathbf{x}$  defined by  $n = \text{len}(\mathbf{x})$ .



Index	Cardboard	Glass	Metal	Paper	Plastic	Trash
Cardboard	67	0	3	11	0	0
Glass	0	69	6	1	23	2
Metal	0	5	68	3	6	0
Paper	5	0	1	108	0	5
Plastic	2	4	2	3	85	1
Trash	1	0	3	4	4	16

FIGURE 6: Confusion matrix (test set) using Softmax classifier global average pooling layer.

The range specifies the number of iterations; we can change the range to ensure a considerable reduction in the corresponding error value.

Image preprocessing, simulation of model, and classification are an end-to-end process described as follows:

- (1) Split the entire dataset into a training set and test set whereby the training set can be denoted as set  $S$  of  $n$  labelled examples:  $S = \{x_i, i = 1, 2, 3 \dots n\}$  where  $n = 2,019$  or  $\sim 80\%$  of the entire dataset, while the test set has 508 images
- (2) Perform data augmentation only on the training set to increase the number of training images. Types of augmentation performed: shearing, zooming, and horizontal flipping of images
- (3) Upon performing data augmentation, the number of the training set is doubled where  $n = 4,038$
- (4) The training set is fed to a pretrained MobileNetV2 CNN, which acts as a feature extractor. Note that the final output layer of the pretrained MobileNetV2 is removed
- (5) The extracted features are used as input to two classifiers, i.e., Softmax and SVM, with 10-fold crossvalidation
- (6) Both models are evaluated using the test set

## 4. Results and Discussion

We have employed TrashNet [32] in the research. The dataset contains 2,527 images of six waste category, namely, cardboard, glass, metal, paper, plastic, and thrash. The images were resized to dimension  $224 \times 224 \times 3$  as required by the MobileNetV2. Data Augmentation such as shear, rotation, zooms, and shifts were used. We split the dataset into training and test sets for comparative analysis.

We train the model using system hardware with 16 GB RAM equipped with NVIDIA RTX 2060 GPU (6GB). To utilize GPU via Python, NVIDIA CUDA, and cuDNN are installed on the system.

*4.1. Performance of MobileNetV2 Using Softmax.* The feature extracted by the pretrained MobileNetV2 obtained from Keras was flattened and input into the newly trained Softmax

TABLE 6: Precision and recall (test set) for MobileNetV2 using Softmax classifier global average pooling layer.

Category	Precision	Recall
Cardboard	0.907	0.840
Glass	0.873	0.683
Metal	0.827	0.817
Paper	0.831	0.908
Plastic	0.712	0.866
Trash	0.640	0.571
Average	<b>0.798</b>	<b>0.781</b>

TABLE 7: Training result (10-fold crossvalidation) for MobileNetV2 with an additional fully connected layer.

Fold	Accuracy	Loss	Precision	Recall
Fold 1	91.337	0.205	0.936	0.889
Fold 2	91.337	0.348	0.936	0.894
Fold 3	88.366	0.387	0.934	0.887
Fold 4	94.059	0.172	0.933	0.885
Fold 5	89.604	0.393	0.931	0.883
Fold 6	89.851	0.295	0.934	0.886
Fold 7	91.089	0.274	0.934	0.887
Fold 8	91.584	0.249	0.932	0.881
Fold 9	91.067	0.255	0.936	0.891
Fold 10	91.067	0.282	0.935	0.888
Average	<b>90.936</b>	<b>0.286</b>	<b>0.934</b>	<b>0.887</b>

classifier. As there are six classes for the waste image dataset, the Softmax classifier layer has 6 nodes. This model will be used as the base model which will be compared with MobileNetV2 using an SVM classifier. Table 1 shows the accuracy, loss, precision, and recall throughout the 10-fold crossvalidation. The 10-fold crossvalidation is a good approximation method that gives unbiased and equal chance to all data points to be used fairly in the training of the model, and it ultimately enhances the accuracy, precision, and recall.

In terms of accuracy, the average accuracy for the 10-fold validation is 88.85%, with an average precision of 0.91 and an average recall of 0.91 as well. Generally, precision and recall

Index	Cardboard	Glass	Metal	Paper	Plastic	Trash
Cardboard	69	0	1	10	0	1
Glass	0	86	5	2	5	3
Metal	0	6	68	1	1	6
Paper	6	0	1	103	0	8
Plastic	1	10	2	3	73	7
Trash	1	0	0	2	0	25

FIGURE 7: Confusion matrix (test set) using MobileNetV2 using Softmax classifier with an additional fully connected layer.

close to 1 indicate that the model performs well in predicting all 6 different categories contained in our dataset. The accuracy of the model on the test set is 79.53%.

Figure 4 presents the confusion matrix of different objects and their classification. The blue cells depict the correct identification of objects related to other objects. Further, the precision and recall are shown in Table 2.

*4.2. Performance of MobileNetV2 Using SVM Classifier.* Similar to the base model presented above, the MobileNetV2 SVM model received the same features extracted by the pre-trained MobileNetV2. Instead of using the Softmax function, SVM was used as the classifier.  $C$  parameter in SVM dictates the amount of regularization performed. Small  $C$  values led to the SVM classifier setting a larger margin when performing classification, while a larger  $C$  value led to a smaller margin. A larger margin (smaller  $C$  values) increased the misclassification rate on the training set but tended to perform better on the test set as it is more generalized. In this study, the default  $C$  value of 1.0 is used.

Table 3 shows the accuracy, loss, precision, and recall throughout the 10-fold crossvalidation for SVM. The SVM model resulted in an average accuracy of 94.28%, average precision of 0.947, and recall of 0.931 which are higher than the 10-fold crossvalidation of the base model (Softmax classifier).

Figure 5 presents the confusion matrix of different objects and their classification. The blue cells depict the correct identification of objects related to other objects. Further, the precision and recall are shown in Table 4.

The MobileNetV2 using the SVM model is then evaluated using the test set, and it provided an accuracy of 83.26%, which is approximately 4% higher than the base model. This is aligned with the result obtained by [31], which had also stated that using the SVM classifier is more accurate when paired together with AlexNet, GoogLeNet, ResNet, and VGG-16 if compared to the Softmax classifier.

*4.3. Performance of MobileNetV2 with Global Average Pooling Using Softmax Classifier.* Based on the results above, it was shown that the SVM model is more accurate than the base model. However, the base model could be further enhanced to increase the accuracy.

A global average pooling layer is added to the model. To illustrate how the global average pooling layer works, given

TABLE 8: Precision and recall (test set) for MobileNetV2 using Softmax classifier with an additional fully connected layer.

Category	Precision	Recall
Cardboard	0.897	0.864
Glass	0.844	0.802
Metal	0.821	0.841
Paper	0.86	0.874
Plastic	0.88	0.753
Trash	0.478	0.786
Average	<b>0.797</b>	<b>0.820</b>

the feature extracted by the pretrained MobileNetV2 model has a dimension of  $7 \times 7 \times 1280$ , after applying global average pooling, the output dimension will be  $1 \times 1280$ .

Table 5 depicts the training result of 10-fold crossvalidation using global average pooling with Softmax classifier with the global average-pooling layer, the model provided an average accuracy of 87.59%, average precision of 0.913, and an average recall of 0.819 from 10-fold crossvalidation. The training accuracy is slightly lower than the base model.

However, the accuracy on the test set is 81.10%, which is higher than the base model. This could be due to the global average-pooling layer which may enable the model to generalize better on the test set.

Figure 6 presents the confusion matrix of different objects and their classification. The blue cells depict the correct identification of objects related to other objects. Further, the precision and recall are shown in Table 6.

*4.4. Performance of MobileNetV2 with Additional Fully Connected Layer.* Using the model with a global average pooling layer, an additional fully connected layer of 64 neurons was added. The new fully connected layer was added so that the model could pick up more complex features, which may be helpful for the waste image dataset.

Table 7 presents the training result of 10-fold crossvalidation for MobileNetV2 with an additional fully connected layer. With 10-fold crossvalidation, MobileNetV2 with an additional fully connected layer provided an average accuracy of 90.93%, average precision of 0.934, and an average recall of 0.887.

Figure 7 presents the confusion matrix of different objects and their classification. The blue cells depict the correct identification of objects related to other objects.

Table 8 describes the precision and recall (test set) for MobileNetV2 using Softmax classifier with an additional fully connected layer. With the newly added fully connected layer, the model achieved an accuracy of 83.46% on the test set, which was higher than the model with a global average pooling layer (81.10%).

## 5. Conclusion

With the significant increase of human living standards, solid waste management, especially recycling, must be emphasised to prevent the adverse environmental problem before it is too late. As the recycling effort requires proper waste segregation by the public, data science, specifically deep learning paired with smart mobile phones, could be a tool to aid the effort in recycling. The development and advancement of CNN and the ability to perform transfer learning are crucial for image classification and model development purposes. Architecture such as VGG16 performs very accurately for image classification. However, the sheer size of a large CNN model, i.e., the number of parameters and operations required, is not feasible for mobile applications.

Although the accuracy of the MobileNetV2 with SVM classifier for waste image classification (83.46%) is lower than the accuracy of the VGG16 model with SVM classifier (97%), using MobileNetV2, which is a more lightweight CNN architecture, is perhaps more practical if the image classification model is to be implemented on mobile devices. Moreover, MobileNetV2 is indeed specifically designed for mobile applications.

In the current research, the feature extracted from MobileNetV2 is directly fed into the SVM classifier. To further improve the MobileNetV2 model SVM classifier, the hyperparameters could be tuned, and perhaps the feature extracted could be fed into multiple layers of the neural network before using SVM so that more complex features or relationships could be extracted. On top of that, the Softmax classifier could be further enhanced by adding more layers as well. The scope of this study is limited to solid waste classification only. We have used cardboard, glass, metal, paper, plastic, and trash but it can be scaled to a large variety of solid waste classification.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors have no conflict of interest in this research.

## Acknowledgments

The authors are grateful to the Taif University Researchers Supporting Project number (TURSP-2020/215), Taif Univer-

sity, Taif, Saudi Arabia. This work is also supported by the Faculty of Computer Science and Information Technology, University of Malaya, under Postgraduate Research Grant (PG035-2016A).

## References

- [1] A. A. Olajire, "The brewing industry and environmental challenges," *Journal of Cleaner Production*, vol. 256, p. 102817, 2020.
- [2] United States Environmental Protection Agency, *Solid Waste and Emergency Response (5306P)*, Washington, DC 20460, 2010, [https://archive.epa.gov/epawaste/nonhaz/municipal/web/pdf/msw\\_2010\\_factsheet.pdf](https://archive.epa.gov/epawaste/nonhaz/municipal/web/pdf/msw_2010_factsheet.pdf).
- [3] U. EPA, *What is integrated solid waste management? Solid Waste and Emergency Response*, Environmental Protection Agency (EPA), United States, 2002.
- [4] US EPA, *Advancing Sustainable Materials Management: Facts and Figures 2013*, United States Environmental Protection Agency, 2015, EPA530-R-15-003.
- [5] H. I. Abdel-Shafy and M. S. M. Mansour, "Solid waste issue: sources, composition, disposal, recycling, and valorization," *Egyptian Journal of Petroleum*, vol. 27, no. 4, pp. 1275–1290, 2018.
- [6] M. A. Kamaruddin, M. S. Yusoff, L. M. Rui, A. M. Isa, M. H. Zawawi, and R. Alrozi, "An overview of municipal solid waste management and landfill leachate treatment: Malaysia and Asian perspectives," *Environmental Science and Pollution Research*, vol. 24, no. 35, pp. 26988–27020, 2017.
- [7] D. Hoornweg and P. Bhada-Tata, *What a Waste: A Global Review of Solid Waste Management*, World Bank, 2014.
- [8] J. R. Jambeck, R. Geyer, C. Wilcox et al., "Plastic waste inputs from land into the ocean," *Science*, vol. 347, no. 6223, pp. 768–771, 2015.
- [9] S. Kumar, J. K. Bhattacharyya, A. N. Vaidya, T. Chakrabarti, S. Devotta, and A. B. Akolkar, "Assessment of the status of municipal solid waste management in metro cities, state capitals, class I cities, and class II towns in India: an insight," *Waste management*, vol. 29, no. 2, pp. 883–895, 2009.
- [10] J. Malinauskaitė, H. Jouhara, D. Czajczyńska et al., "Municipal solid waste management and waste-to-energy in the context of a circular economy and energy recycling in Europe," *Energy*, vol. 141, pp. 2013–2044, 2017.
- [11] E. I. Ugwu, A. C. Ekeleme, P. O. Awoyera, H. O. Ozioko, and U. Osinachi, "Effect of municipal solid waste contamination on some geotechnical properties of soil," *Journal of Materials and Environmental Sciences*, vol. 9, no. 2, pp. 585–590, 2018.
- [12] R. G. Alfaia, A. M. Costa, and J. C. Campos, "Municipal solid waste in Brazil: a review," *Waste Management and Research*, vol. 35, no. 12, pp. 1195–1209, 2017.
- [13] R. M. Rodzi, Z. M. Nopiah, and N. E. A. Basri, "Analysis of solid waste generation and composition in Malaysia TVET campus," *International Journal of Integrated Engineering*, vol. 11, no. 2, 2019.
- [14] B. Esmaeilian, B. Wang, K. Lewis, F. Duarte, C. Ratti, and S. Behdad, "The future of waste management in smart and sustainable cities: a review and concept paper," *Waste Management*, vol. 81, pp. 177–195, 2018.
- [15] W. A. Qazi, M. F. M. Abushammala, and M. H. Azam, "Multi-criteria decision analysis of waste-to-energy technologies for municipal solid waste management in Sultanate of Oman,"

- Waste Management & Research*, vol. 36, no. 7, pp. 594–605, 2018.
- [16] N. Pour, P. A. Webley, and P. J. Cook, “Potential for using municipal solid waste as a resource for bioenergy with carbon capture and storage (BECCS),” *International Journal of Greenhouse Gas Control*, vol. 68, pp. 1–15, 2018.
- [17] I. R. Istrate, D. Iribarren, J. L. Gálvez-Martos, and J. Dufour, “Review of life-cycle environmental consequences of waste-to-energy solutions on the municipal solid waste management system,” *Resources, Conservation and Recycling*, vol. 157, p. 104778, 2020.
- [18] P. Yadav and S. R. Samadder, “A critical review of the life cycle assessment studies on solid waste management in Asian countries,” *Journal of Cleaner Production*, vol. 185, pp. 492–515, 2018.
- [19] G. Bertanza, E. Ziliani, and L. Menoni, “Techno-economic performance indicators of municipal solid waste collection strategies,” *Waste Management*, vol. 74, pp. 86–97, 2018.
- [20] M. Kannangara, R. Dua, L. Ahmadi, and F. Bensebaa, “Modeling and prediction of regional municipal solid waste generation and diversion in Canada using machine learning approaches,” *Waste Management*, vol. 74, pp. 3–15, 2018.
- [21] C. E. Kontokosta, B. Hong, N. E. Johnson, and D. Starobin, “Using machine learning and small area estimation to predict building-level municipal solid waste generation in cities,” *Computers, Environment and Urban Systems*, vol. 70, pp. 151–162, 2018.
- [22] K. Pardini, J. J. Rodrigues, S. A. Kozlov, N. Kumar, and V. Furtado, “IoT-based solid waste management solutions: a survey,” *Journal of Sensor and Actuator Networks*, vol. 8, no. 1, p. 5, 2019.
- [23] O. Adedeji and Z. Wang, “Intelligent waste classification system using deep learning convolutional neural network,” in *Procedia Manufacturing*, vol. 35, pp. 607–612, 2019.
- [24] M. Abbasi and A. El Hanandeh, “Forecasting municipal solid waste generation using artificial intelligence modelling approaches,” *Waste Management*, vol. 56, pp. 13–22, 2016.
- [25] J. N. Kalshetty, A. Anil, H. Jain, S. Beli, and A. K. Singh, “Smart monitoring for waste management using internet of things,” in *International Conference on Intelligent Data Communication Technologies and Internet of Things*, pp. 96–103, Cham, 2020.
- [26] A. Anil, H. Jain, S. Beli, A. K. Singh, and J. N. Kalashetty, “Smart monitoring for waste management using IoT,” *International Journal of Engineering and Advanced Technology*, vol. 9, no. 1, pp. 7502–7505, 2019.
- [27] S. Sudha, M. Vidhyalakshmi, K. Pavithra, K. Sangeetha, and V. Swaathi, “An automatic classification method for environment: friendly waste segregation using deep learning,” in *2016 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR)*, Chennai, India, 2016.
- [28] M. Yang and G. Thung, *Classification of Trash for Recyclability Status*, 2016, CS229Project Rep..
- [29] M. Satvilkar and P. N. Cosgrave, *Image Based Trash Classification Using Machine Learning Algorithms for Recyclability Status*, Doctoral dissertation, National College of Ireland, Dublin, 2018.
- [30] A. Arayakandy, “Design and development of classification model for recyclability status of trash using SVM,” *International Journal for Research in Applied Science and Engineering Technology*, vol. 7, no. 3, pp. 2146–2150, 2019.
- [31] U. Özkaya and L. Seyfi, “Fine-tuning models comparisons on garbage classification for recyclability,” 2018, <https://arxiv.org/abs/1908.04393>.
- [32] T. Kennedy, “OscarNet: using transfer learning to classify disposable waste,” *CS230 Report: Deep Learning*, Stanford University, CA, Winter, 2018.
- [33] A. G. Howard, M. Zhu, B. Chen et al., “MobileNets,” 2017, <https://arxiv.org/abs/1704.04861>.
- [34] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, “MobileNetV2: inverted residuals and linear bottlenecks,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Salt Lake City, Utah, state of Utah, USA, 2018, Calvin L. Rampton Salt Palace Convention Center the week of June 18-22.
- [35] M. Hollemans, “MobileNet version 2,” *machine think BLOG*, 2018.

## Research Article

# A Denoising Autoencoder-Based Bearing Fault Diagnosis System for Time-Domain Vibration Signals

Yi Gu , Jiawei Cao, Xin Song, and Jian Yao

*School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi, Jiangsu 214122, China*

Correspondence should be addressed to Yi Gu; [8202101437@jiangnan.edu.cn](mailto:8202101437@jiangnan.edu.cn)

Received 6 April 2021; Revised 27 April 2021; Accepted 4 May 2021; Published 15 May 2021

Academic Editor: Shan Zhong

Copyright © 2021 Yi Gu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The condition monitoring of rotating machinery is always a focus of intelligent fault diagnosis. In view of the traditional methods' excessive dependence on prior knowledge to manually extract features, their limited capacity to learn complex nonlinear relations in fault signals and the mixing of the collected signals with environmental noise in the course of the work of rotating machines, this article proposes a novel approach for detecting the bearing fault, which is based on deep learning. To effectively detect, locate, and identify faults in rolling bearings, a stacked noise reduction autoencoder is utilized for abstracting characteristic from the original vibration of signals, and then, the characteristic is provided as input for backpropagation (BP) network classifier. The results output by this classifier represent different fault categories. Experimental results obtained on rolling bearing datasets show that this method can be used to effectively diagnose bearing faults based on original time-domain signals.

## 1. Introduction

Rolling bearings are the modules that have frequent usage in mechanical equipment and are also among the most vulnerable components. When a fault occurs, it often results in enormous losses; consequently, bearing fault diagnosis technology is receiving increasing attention. Statistics show that around 30% of machinery failures arise from bearings [1, 2]. To comprehensively detect the health status of rotating machinery, a condition monitoring system is used for data collection. After a long period of operation, such a machine will have generated a large amount of real-time data. We can obtain the features that are extracted from the monitored data through constructing feature engineering. These data can be used to establish a classifier that different bearing faults can be identified [3, 4]. Therefore, detecting bearing faults is essentially the task of classification [5–7], and the health status of bearings can be identified through such artificial intelligence techniques as deep learning models [8–10], support vector machines (SVMs) [11], neural networks (NNs) [12, 13], extreme learning machines (ELMs) [14], and autoencoders (AEs) [15–17]. In particular, in recent

years, deep learning technology has been widely used in this field and achieved good results [18, 19].

Most of the common methods used to detect bearing faults at present are supervised algorithms where labor label cost is pretty high. Because of the powerful function of hidden layers [20–22], the algorithm for training the classifier that detects bearing faults is NNs commonly. However, high-performance neural network classifiers also rely on numerous good-quality labeled data. But overfitting of NNs may occur when the samples used for training is noisy, or the amount is not enough, or the test distribution cannot be covered. Thus, the ability of NNs to generalize becomes poor particularly for complicated classification problems. Diagnosing bearing faults is challenging because the subject to be solved is a sophisticated mechanical instrument [23, 24]. Hence, lots of approaches have been put forward to figure this issue out, like senior techniques of processing the signals that are used for analyzing the vibration of signal to enable the extraction of helpful bearing fault features. But these advanced techniques require practitioners to have a deep knowledge reserve of vibration signal and the whole mechanical systems. Moreover, because it is difficult to obtain the

professional knowledge related to this, these methods are not universal, meaning that they are not smart enough compared to machine learning. As an alternative and with the power of deep learning, deep neural networks (DNNs) are proposed to achieve unsupervised learning of characteristic. In recent years, considering the impressive achievements in the area of image recognition, deep learning has become the most popular and promising research methods. Among them, deep AE (autoencoder) network structures, as one of the representatives of unsupervised learning for detecting bearing faults, have become a common solution. In addition, as a well-known unsupervised feature learning method, denoising self-encoding is widely used for the realization of various tasks because it can be used to learn more robust feature expressions for input signals, that is, it has a strong generalization ability [25–27].

## 2. Brief Introduction to DNNs

DNN is usually a nonlinear converter with multilayer structures. In general, it contains an input layer, an output layer, and several layers in the middle. The parameter values of each layer in DNN can be learned by the training mechanism of the neural network, as Hinton et al. said [10]. By comparison with smaller NNs, DNNs can express more information. With the power of deep networks, we propose to use a SAE-based (stacked autoencoder) method to automatically classify bearing faults through feature selection without any manual intervention. This method consists of three steps. Step one: use unsupervised learning to train the entire model gradually. Step two: use supervised learning to fine-tune this network with the training data which contain true labels. Step three: use the network to perform feature extraction and classification of the bearing data after completing the training of this network.

A DAE (denoising autoencoder) [25] is a variant of an AE obtained by adding denoising technology to an AE. Traditionally, an AE comprises an encoder and a decoder. Besides, AE is also one of the representative methods of unsupervised learning. The principle that AE can achieve unsupervised feature extraction is through continuously minimizing the deviation between the input and the reconstructed data. For the sake of enhancing the generalization performance of the entire model, we add some noise to the input layer. This approach also effectively avoids overfitting in the training of the network. Usually, a DAE is a symmetrical neural network with an odd number of layers, and its goal is to make the output and input data as consistent as possible. The purpose of the encoder of DAE is to condense the input data without losing important information as much as possible and get a concise representation of the input data. The diagram of a DAE is as seen in Figure 1.

$$f_{\{W^1, b^1\}}(x) = s_f(W^1 x + b^1), \quad (1)$$

$$g_{\{W^2, b^2\}}(x) = s_g(W^2 x + b^2). \quad (2)$$

In Figure 1, the encoder is made up by an input layer and

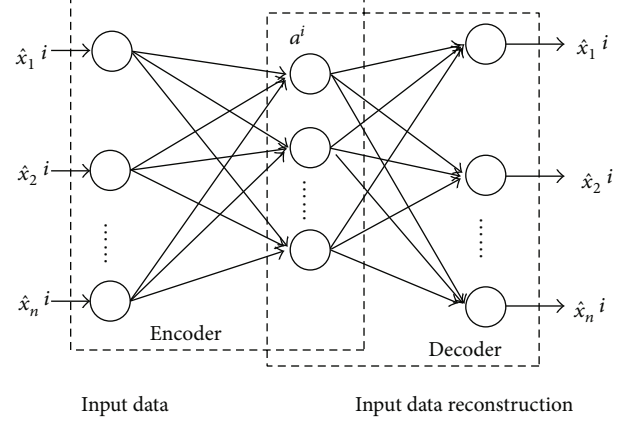


FIGURE 1: Structure graph of an AE.

parts of the layers near the input, and the decoder is composed by an output layer and parts of the layers near the output. The process that the encoder performs processing can be expressed by Equation (1), and Equation (2) represents the processing performed by the decoder, where  $s_f$  and  $s_g$ , respectively, denote as the activation functions of the coding process and the decoding process. For these activation functions, the nonlinear sigmoid function is commonly used. Suppose a dataset is represented as  $X = \{x^i\}_{i=1}^M$ , where  $M$  represents the size of  $X$ . The expression of the sample  $x^i$  in  $X$  obtained by the encoder is shown as  $a^i$ . Furthermore,  $a^i$  is also a concise expression of the input.

$$a^i = f_{\{W^1, b^1\}}(x^i). \quad (3)$$

Suppose the reconstructed error is represented as  $x \wedge^i$ . The connection between  $x \wedge^i$  and  $a^i$  is expressed as

$$x \wedge^i = g_{\{W^2, b^2\}}(a^i). \quad (4)$$

The parameters of DAE are updated iteratively during the training phase. The objective function of DAE is as follows:

$$J(W, b) = \|x - g(f(\tilde{x}))\|_2^2, \quad (5)$$

$$\tilde{x} = \text{dropout}(x).$$

In Equation (5),  $\tilde{x}$  represents the input  $x$  after adding some noise. In DAE, it is implemented through dropout technology. Therefore, DAE must be able to recover  $x$  from this pollution rather than just using the data with noise. By this means, the inner distribution of the generated data structure can be implicitly captured by the encoding function  $f(\bullet)$  and the decoding function  $g(\bullet)$  of DAE in common, thus enabling the extraction of robust features. DAE-based automatic detecting for bearing faults method.

We propose a novel automatic detecting for bearing faults method in this article. This method is based on DAE. The breakdown characteristic from primitive vibration signals of rotating equipment like bearings can be mined by this

method. By adaptively selecting these representative features, this method can automatically classify the condition category, which is the health level, of the machinery on these breakdown features. We measure the primitive signals of vibration in the time domain. The analysis of converted spectral signals of time domain and frequency domain is opposite, where frequency domain applies the fast Fourier transform (FFT). The temporal waveform determines the analysis in the time domain, and such a time-domain signal provides relatively vague information on the health of the rotating machinery. Therefore, it is very challenging to classify these original time-domain signals directly.

The proposed method contains four steps as seen in Figure 2.

Step one: acquire time-domain data from rotating equipment with various states of health. Assume that the training dataset composed of these time-domain data is denoted as  $\{x^i, d^i\}_{i=1}^M$ , where  $x^i$  represents the  $i$ -th sample in the training-phase,  $d^i$  represents the state label of health related to  $x^i$ , and  $M$  represents the amount of time-domain data.

Step two: construct a DNN which consists of several hidden layers. The amount of neurons in the first layer is identical to the dimensionality of the time-domain data  $x^i$ . Next, pretrain the DNN on each layer through SAE with the unlabeled dataset  $x = \{x^i\}_{i=1}^M$ . In the pretraining, the required number of AE is the same as that of the amount of hidden layers in DNN. The specific method of pretraining layer by layer is that firstly take the first three layers of DNN, where it consists of an input and two hidden layers, as the three-layer structure of AE1. The correspondence between above two is that the input layer of DNN is also the input layer of AE1, the hidden layer 1 of DNN is the hidden layer of AE1 and the hidden layer 2 of DNN is the output layer of AE1. The dataset of training AE1 is the unlabeled dataset  $x$ , which is also regarded as the output of AE1. After completing the training for AE1, use the parameters  $\{W^1, b^1\}$  of AE1 to finish the initialization of the parameters of hidden layer 1 of DNN. Let  $a^1$  be the encoding vector which is from time-domain data of the time rotating machinery by AE1. Then, take  $a^1$  as the input data and objective output of AE2 for training and finish the initialization of the parameters of hidden layer 2 of DNN via  $\{W^2, b^2\}$ , and obtain  $a^2$  accordingly. Finally, repeat the above steps sequentially till AEN is obtained and the time-domain data is encoded as  $a^N$ . By using the above method, we complete the pretraining of the whole of the middle layers of DNN.

Step three: the size of the last layer of DNN, which performs classification, should be consistent with the amount of different states of health that machine may have. Then, by making the reconstruction value from the original time-domain data as close to the given health status label as possible, the parameters of DNN are fine-tuned. The method to optimize the entire network is the backpropagation (BP) algorithms.

Step four: apply the trained DNN to faults diagnose of rotating equipment.

In the approach of this paper, DNN learns multiple nonlinear transformations through a pretraining process to dis-

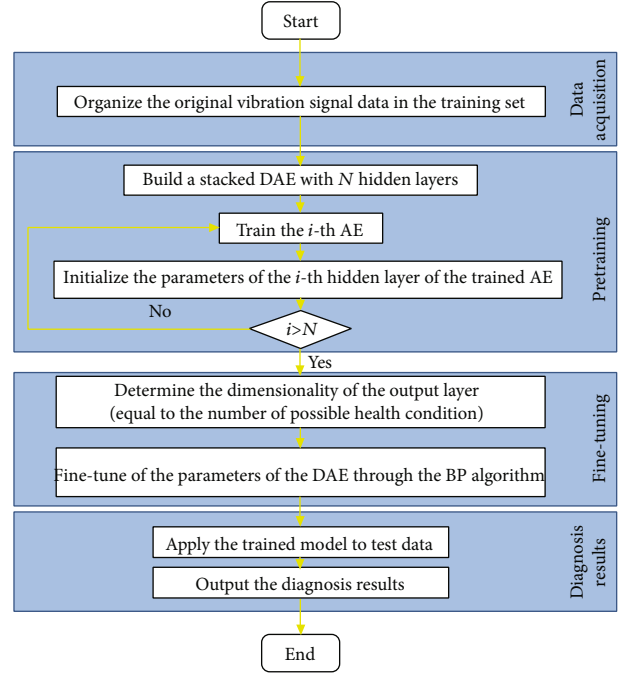


FIGURE 2: Framework of the method in this paper.

cover the main changes in the time-domain signal characteristics. In this way, DAE can extract more robust features. After that, the fine-tuning process can help DNN mine discriminative information in the vibration signal data. By these means, the well-trained DNN is capable of automatically capturing the fundamental features from the time-domain data and set up a complex nonlinear relationship between the bearing status labels and time-domain data. Thus, this method can be used to realize effective attribute selection and rapid smart detecting of rotating bearing breakdown features.

### 3. Fault Diagnosis Experiment

Rolling bearings are the core parts in rotating machinery. The performance, service life, and reliability of this type of machine usually depend on the health of these components. However, because of their harsh working environment, these components are susceptible to various damages, resulting in malfunctions and causing serious economic damage. Several diagnostic situations of rolling bearings will be considered to verify the proposed method in this part.

**3.1. Introduction to the Training Data.** The raw data of rotating bearing were gathered from a system of machine which is driven by a motor. The approach of acquisition is to utilize a specific frequency to complete the sampling, where the frequency is 12 kHz. In general, there are four ways of loading in these bearing data. Thus, the data were obtained under the following four experimental conditions: (a) normal state (N), (b) outer race fault (OF), (c) inner race fault (IF), and (d) roller fault (RF). Meanwhile, when constructing the raw data, the diameter of the failed rotating bearing is also considered into the driving termination engine. In this method,

TABLE 1: Descriptions of the training data.

Datasets	Load (hp)	The amount of samples	Fault type	Fault size	Classification label
A B C D E	0 1 2 3 0-3	600/600/600/1800	N	0''	1
		600/600/600/1800	RF	0.007''	2
		600/600/600/1800	RF	0.014''	3
		600/600/600/1800	RF	0.021''	4
		600/600/600/1800	IF	0.007''	5
		600/600/600/1800	IF	0.014''	6
		600/600/600/1800	IF	0.021''	7
		600/600/600/1800	OF	0.007''	8
		600/600/600/1800	OF	0.014''	9
		600/600/600/1800	OF	0.021''	10

TABLE 2: Diagnosis results obtained on different bearing datasets.

Dataset	Metric	ELM	SVM	BPNN1 (500)	BPNN2 (500-500)	BPNN3 (600-200-100)	SAE+BP	Proposed method
A	Acc-mean	11.23	77.07	83.97	85.96	78.01	86.32	92.29
	Acc-std	0.82	0.57	0.37	0.56	0.49	0.39	0.19
B	Acc-mean	11.6	74.47	85.77	88.58	79.87	87.14	93.32
	Acc-std	0.86	0.61	0.47	0.36	0.39	0.30	0.33
C	Acc-mean	11.70	72.69	89.28	89.24	77.57	88.92	93.76
	Acc-std	0.86	0.50	0.37	0.41	0.30	0.28	0.25
D	Acc-mean	11.37	72.45	87.67	89.57	81.13	88.70	94.06
	Acc-std	0.52	0.00	0.43	0.34	0.35	0.43	0.21
E	Acc-mean	13.10	93.34	96.11	97.84	93.55	97.79	99.04
	Acc-std	1.19	0.28	0.13	0.10	0.13	0.12	0.05

there are three sizes of the fault: 0.007 inches, 0.014 inches, and 0.021 inches. In this article, five datasets, respectively, named A, B, C, D, and E, are used for checking the classification capability of this method. The detailed description of these five datasets is seen in Table 1. Furthermore, the health states of bearing data are divided into ten different types of labels, named 1 to 10. Therefore, A, B, C, and D in the datasets cover these condition type labels with the loading of 0, 1, 2, and 3 hp, respectively. Each state of health involves 600 signals, and there are 2048 data points for every single signal.

**3.2. Experimental Setup.** The designed one side of DNN consists of four layers. The size of the first layer of DNN is identical to the feature dimension of the training data, which is 2048. As for the size of the hidden layers in the middle of the encoder, we uniformly set them to 500. The size of the last layer of the encoder depends on the categories of possible bearing states of health. Specifically, the shape of the entire DNN is 2048-500-500-10-500-500-2048, where the structure of encoder and decoder is symmetric. We use rectified linear units (ReLUs) in all encoder/decoder pairs except the first (which needs to reconstruct the input data, which may have both positive and negative values) and the last (to ensure that the final encoding retains complete information). For these experiments, during greedy layerwise pretraining, the initial-

ization of the weights is set to a random number with a standard deviation of 0.01 derived from a zero-mean Gaussian distribution. Each layer was pretrained for 500 iterations with a dropout rate of 20% and a learning rate of 0.001. We select 50% of the data at random to pretrain the network, and at the same time, these samples are also utilized to optimize the entire network through fine-tuning the parameters. After that, the rest of the data (another 50% data) are used for testing the classification capability. The fine-tuning process was performed by removing the decoder from the pretrained DAE and adding a soft-max layer for BP network, and we set the maximum value of fine-tuning epochs to 500. Table 2 shows the diagnostic results of ten tests performed on each dataset in A to E using the proposed method. We are aware from the experimental results of these test sets that the accuracy of diagnose for bearing fault data is greater than 92%, and some even reach to 99%, which signifies that our method is capable of well distinguishing the status of health of ten types of bearings.

**3.3. Comparison with Traditional Methods.** Backpropagation neural networks (BPNNs) and SVMs have been widely used in fault detection for rotating equipment. In this part, we compare our method with a multiple layer architecture network, that is, BPNN, whose structure is 2048-500-500-10,



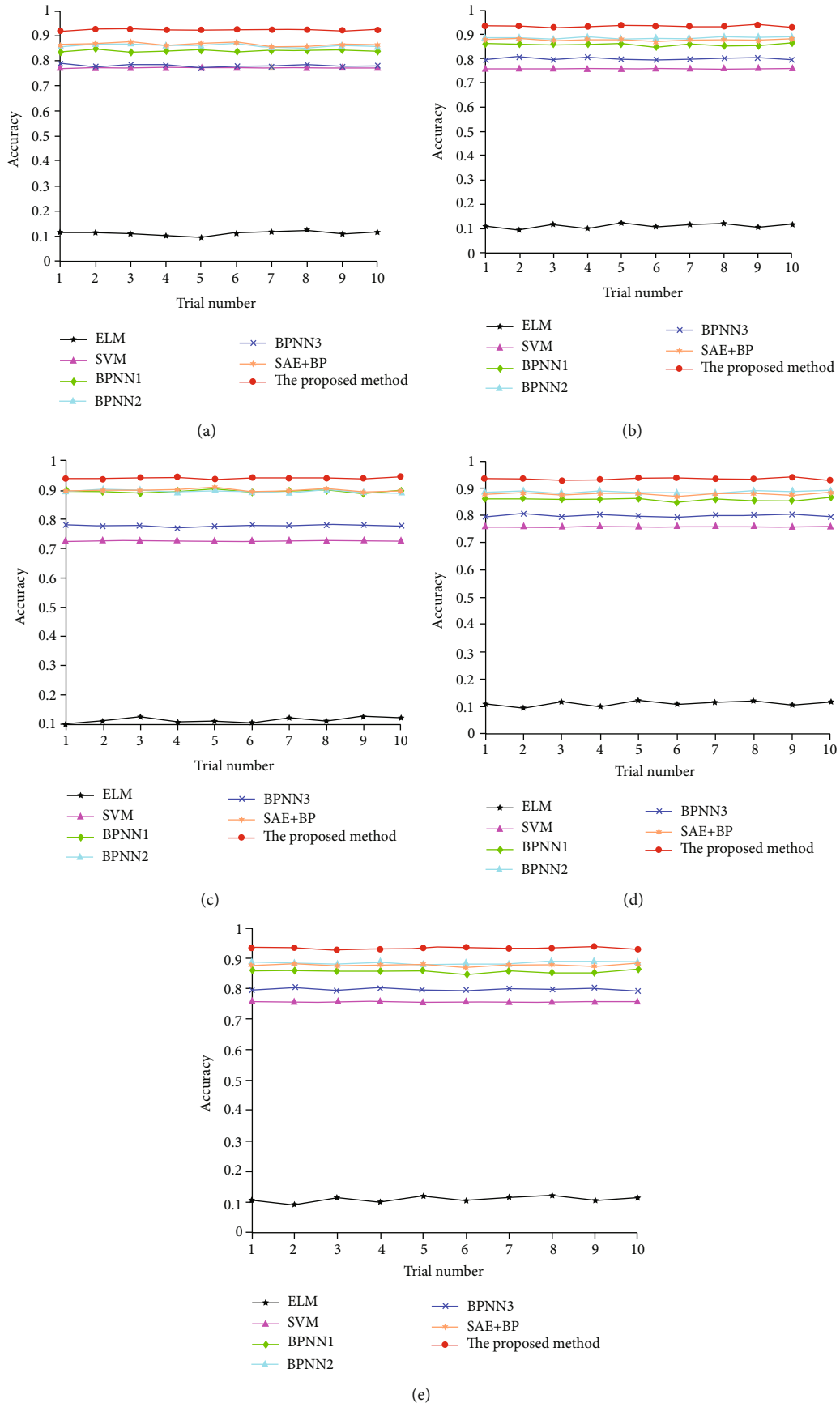


FIGURE 3: Diagnosis results from 10 trials on different bearing datasets: (a) A, (b) B, (c) C, (d) D, and (e) E.

another multi-hidden-layer BPNN used in a previous study, whose network structure is 2048-600-200-100-10, a neural network with only one hidden layer, whose network structure is 2048-500-10, an SVM-based model, an ELM, and a traditional SAE-based model with the same architecture used in the proposed method. A comparison of the diagnostic accuracy of the different methods across 10 experiments is seen in Figure 3.

The test results demonstrate that the method with the most accurate detection and the lowest volatility is our proposed method, and they also show a greater advantage in stability. Compared with the proposed method, the SVM has relatively stable classification performance and slightly greater volatility. What is more, compared with other methods, its diagnostic performance is relatively stable. Compared with our method or SVM, only one hidden layer BPNN shows worse recognition precision and greater volatility. In addition, the performance of multilayer BPNN2 is only better than that of smaller BPNN1, but the diagnostic results of the multi-hidden-layer BPNN3, with the deepest structure, are poorer. This shows that the classification accuracy of a BPNN cannot always be improved simply by increasing the number of network layers; instead, a reasonable network structure must be selected. In addition, we can see that the results obtained from the traditional SAE-based method are similar to the results of BPNN1, which has the same hidden layer structure. This indicates that even if a DNN is used for pretraining, it is not necessarily possible to effectively extract the useful features of the original time-domain data. In addition, ELM achieves the worst results in all experiments, indicating that this classification method cannot process the original time-domain data effectively. We calculated the mean diagnostic accuracies and standard deviations of more than 10 tests for quantitative comparisons. Table 2 details the experimental results.

It is worth noting that, in each trial, the same training dataset is used for checking the performance of each comparison approach. From the perspective of average classification accuracy, our proposed method can achieve the best diagnostic accuracy, while the ELM has the worst diagnostic accuracy. Moreover, our method also has the best performance in terms of classification stability. The results indicate that our method in this article is able to effectively and stably identify the type of bearing failure and the severity of damage to rotating machinery. In comparison, the results obtained by BPNN in the table are relatively poor in both average accuracy and standard deviation, which shows that our method is better than BPNN. In particular, dataset E contains a large number of samples representing the states of health of bearing data with ten different loadings, and the detecting precision of the method proposed in this paper on this dataset indicates that this approach can detect bearing breakdowns independently from load fluctuations.

The above results show that, in comparison with the conventional diagnosis approaches, our proposed approach of combining unsupervised DAE pretraining with the supervised BP algorithm has significant advantages. In this method, the DAE is trained with the reconstruction errors generated from multidimensional temporal data to explore

the relationship between time-domain data. Then, the decoder is replaced with a soft-max layer for health condition monitoring and fault diagnosis. Therefore, compared with the traditional recognition methods based on time-domain signal analysis, the proposed method can obtain higher recognition accuracy by means of a reasonable network structure and the introduction of dropout technology in the pretraining process. The characteristic structure of a time-domain vibration signal is fuzzy, and its regularity is poor. Therefore, the method of combining traditional time-domain features with an SVM classifier results in a low average diagnostic accuracy. However, a DAE is capable of deep learning from the original time-domain data, mining hidden information and identifying differences. Therefore, the DAE-based classification method proposed in this paper can obtain high recognition accuracy in repeated trials and maintain stable performance. Because of the challenges presented by the noise in the original time-domain signals, although the traditional SAE-based method also relies on a deep structure for network pretraining, good classification performance cannot be obtained. In addition, because the structure of a BPNN is relatively shallow and its feature learning ability is limited, its diagnostic accuracy is not high. The BP algorithm based on a multi-hidden-layer BPNN results in unstable network performance and poor generalization ability.

#### 4. Conclusions

We come up with a smart method in this article which is on the basis of DAE to diagnose faults in rotating machinery. The feasibility of this approach for breakdown detecting of rolling bearings is verified on five test datasets. These datasets contain a large number of samples that represent the health of different bearings under various operating conditions. It demonstrates that based on the diagnosis outcome obtained from these datasets this method can adaptively mine fault features from time-domain data, so as to successfully solve a variety of detecting problems and efficiently implement the classification of the condition of the machine. Throughout the introduction of data dropout, the generalization ability and anti-interference ability of the model are improved. Since the fault features are automatically extracted, compared with traditional BPNN-based methods, our proposed method can avoid the dependence on artificial label and without the expert-level understanding of mechanical signal and diagnosis so that it can be easily used in practical applications.

#### Data Availability

The labeled dataset used to support the findings of this study are available from the corresponding author upon request.

#### Conflicts of Interest

The authors declare no conflicts of interest.

## Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 61772241 and in part by the 2018 Six Talent Peaks Project in Jiangsu Province under Grant XYDXX-127.

## References

- [1] T. Junbo, L. Weining, A. Juneng, and W. Xueqian, "Fault diagnosis method study in roller bearing based on wavelet transform and stacked auto-encoder," in *the 27th Chinese Control and Decision Conference (2015 CCDC)*, Qingdao, China, 2015.
- [2] T. D. Bui and Guangyi Chen, "Translation-invariant denoising using multiwavelets," *IEEE Transactions on Signal Processing*, vol. 46, no. 12, pp. 3414–3420, 1998.
- [3] Y. Bengio and O. Delalleau, "On the expressive power of deep architectures," in *Algorithmic Learning Theory. ALT 2011*, J. Kivinen, C. Szepesvári, E. Ukkonen, and T. Zeugmann, Eds., vol. 6925 of Lecture Notes in Computer Science, pp. 18–36, Springer, Berlin, Heidelberg, 2011.
- [4] G. E. Hinton, S. Osindero, and Y. W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.
- [5] H. Liu, J. Zhou, Y. Zheng, W. Jiang, and Y. Zhang, "Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders," *ISA Transactions*, vol. 77, pp. 167–178, 2018.
- [6] F. Pacheco, M. Cerrada, R. V. Sánchez, D. Cabrera, C. Li, and J. Valente de Oliveira, "Attribute clustering using rough set theory for feature selection in fault severity classification of rotating machinery," *Expert Systems with Applications*, vol. 71, pp. 69–86, 2017.
- [7] H. Shao, H. Jiang, H. Zhao, and F. Wang, "A novel deep auto-encoder feature learning method for rotating machinery fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 95, no. 95, pp. 187–204, 2017.
- [8] M. He and D. He, "Deep learning based approach for bearing fault diagnosis," *IEEE Transactions on Industry Applications*, vol. 53, no. 3, pp. 3057–3065, 2017.
- [9] Y. Wang, M. Liu, and Z. Bao, "Deep learning neural network for power system fault diagnosis," in *2016 35th Chinese Control Conference (CCC)*, pp. 6678–6683, Chengdu, China, 2016.
- [10] G. E. Hinton and R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
- [11] P. Chen, L. Yuan, Y. He, and S. Luo, "An improved SVM classifier based on double chains quantum genetic algorithm and its application in analogue circuit diagnosis," *Neurocomputing*, vol. 211, pp. 202–211, 2016.
- [12] B. Bessam, A. Menacer, M. Boumehraz, and H. Cherif, "Detection of broken rotor bar faults in induction motor at low load using neural network," *ISA Transactions*, vol. 64, pp. 241–246, 2016.
- [13] M. A. Atoui, S. Verron, and A. Kobi, "Fault detection with conditional Gaussian network," *Engineering Applications of Artificial Intelligence*, vol. 45, pp. 473–481, 2015.
- [14] M. Luo, C. Li, X. Zhang, R. Li, and X. An, "Compound feature selection and parameter optimization of ELM for fault diagnosis of rolling element bearings," *ISA Transactions*, vol. 65, pp. 556–566, 2016.
- [15] C. Lu, Z. Wang, W. Qin, and J. Ma, "Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification," *Signal Processing*, vol. 130, no. 130, pp. 377–388, 2017.
- [16] F. Jia, Y. Lei, J. Lin, X. Zhou, and N. Lu, "Deep neural networks: a promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data," *Mechanical Systems and Signal Processing*, vol. 72–73, pp. 303–315, 2016.
- [17] J. Sun, C. Yan, and J. Wen, "Intelligent bearing fault diagnosis method combining compressed data acquisition and deep learning," *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 1, pp. 185–195, 2018.
- [18] M. Sohaib and J. Kim, "Reliable fault diagnosis of rotary machine bearings using a stacked sparse autoencoder-based deep neural network," *Shock and Vibration*, vol. 2018, Article ID 2919637, 11 pages, 2018.
- [19] W. Sun, S. Shao, R. Zhao, R. Yan, X. Zhang, and X. Chen, "A sparse auto-encoder-based deep neural network approach for induction motor faults classification," *Measurement*, vol. 89, pp. 171–178, 2016.
- [20] J. Lee, F. Wu, W. Zhao, M. Ghaffari, L. Liao, and D. Siegel, "Prognostics and health management design for rotary machinery systems—reviews, methodology and applications," *Mechanical Systems and Signal Processing*, vol. 42, no. 1–2, pp. 314–334, 2014.
- [21] K. Worden, W. J. Staszewski, and J. Hensman, "Natural computing for mechanical systems research: a tutorial overview," *Mechanical Systems and Signal Processing*, vol. 25, no. 1, pp. 4–111, 2011.
- [22] B. Samanta, "Artificial neural networks and genetic algorithms for gear fault detection," *Mechanical Systems and Signal Processing*, vol. 18, no. 5, pp. 1273–1282, 2004.
- [23] B. Samanta and C. Nataraj, "Use of particle swarm optimization for machinery fault detection," *Engineering Applications of Artificial Intelligence*, vol. 22, no. 2, pp. 308–316, 2009.
- [24] V. T. Tran, B. Yang, M. Oh, and A. C. C. Tan, "Fault diagnosis of induction motor based on decision trees and adaptive neuro-fuzzy inference," *Expert Systems with Applications*, vol. 36, no. 2, pp. 1840–1849, 2009.
- [25] P. Vincent, H. Larochelle, I. Lajoie, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion," *Journal of Machine Learning Research*, vol. 11, pp. 3371–3408, 2010.
- [26] Q. V. Le, M. Ranzato, R. Monga et al., "Building high-level features using large scale unsupervised learning," in *International conference on machine learning*, pp. 507–514, Vancouver, BC, Canada, 2012.
- [27] J. Xie, R. Girshick, and A. Farhadi, "Unsupervised deep embedding for clustering analysis," in *International conference on machine learning*, pp. 478–487, New York, USA, 2016.

## Research Article

# Chinese Personal Name Disambiguation Based on Clustering

Chao Fan <sup>1,2</sup> and Yu Li<sup>1,2</sup>

<sup>1</sup>The School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi 214122, China

<sup>2</sup>Jiangsu Key Laboratory of Media Design and Software Technology, Jiangnan University, Wuxi 214122, China

Correspondence should be addressed to Chao Fan; [fanchao@jiangnan.edu.cn](mailto:fanchao@jiangnan.edu.cn)

Received 9 April 2021; Revised 26 April 2021; Accepted 29 April 2021; Published 15 May 2021

Academic Editor: Shan Zhong

Copyright © 2021 Chao Fan and Yu Li. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Personal name disambiguation is a significant issue in natural language processing, which is the basis for many tasks in automatic information processing. This research explores the Chinese personal name disambiguation based on clustering technique. Preprocessing is applied to transform raw corpus into standardized format at the beginning. And then, Chinese word segmentation, part-of-speech tagging, and named entity recognition are accomplished by lexical analysis. Furthermore, we make an effort to extract features that can better disambiguate Chinese personal names. Some rules for identifying target personal names are created to improve the experimental effect. Additionally, many calculation methods of feature weights are implemented such as bool weight, absolute frequency weight, tf-idf weight, and entropy weight. As for clustering algorithm, an agglomerative hierarchical clustering is selected by comparison with other clustering methods. Finally, a labeling approach is employed to bring forward feature words that can represent each cluster. The experiment achieves a good result for five groups of Chinese personal names.

## 1. Introduction

The ambiguity of named entities is a prevalent phenomenon in natural language. There is considerable ambiguity about the personal name in the texts or the web pages, especially in the Chinese dataset. The Chinese personal name “Gao Jun (高军)” has a total of 51 items in the Baidu Encyclopedia. Eliminating the ambiguity of such personal name is beneficial to many tasks like information retrieval and data summarization. Take searching a person name on the Internet for example, documents of different person entities with the same name can be found by search engine. It is necessary to divide the documents into clusters automatically and secure the key information of each cluster. This research focuses on this task of importance and attempts to solve the problem by unsupervised approaches.

Chinese personal name disambiguation involves distinguishing between people with an ambiguous name in Chinese corpus. Initially, documents with the html format in raw corpus are processed into plain texts. Then, the lexical analysis of documents is performed, including segmentation, part-of-speech tagging (POS tagging), and named entity recognition (NER). Feature selection is enforced according to

the result of lexical analysis. In order to acquire better accuracy of personal name recognition, some rules of personal name extension are proposed for target names to be disambiguated. For instance, the family name and first name of a target name may be separated in some situation due to the segmentation errors. We merge them into one complete personal name with the purpose of reducing the number of discarded documents. Further, an agglomerative hierarchical clustering algorithm is adopted to discover different clusters containing the same personal name. Finally, the label of each cluster is given by scoring the weight of each feature word in cluster. The feature words chosen as the cluster label can represent the person entity with significant information.

The rest of this article is arranged in the following parts. Related work of this task is introduced in Section 2. Research framework and methodology are elaborated in Section 3. Section 4 gives the experimental results and some discussions. Conclusion and future work are discussed in Section 5.

## 2. Related Work

The personal name disambiguation task is similar to the word sense disambiguation (WSD). Both of them pursue

the goal of resolving the ambiguity in natural language understanding. Nevertheless, there is a big difference between two tasks. The number of person entities with an ambiguous personal name is usually unknown for the name disambiguation task, which is contrary to WSD. Hence, personal name disambiguation is often implemented with an unsupervised clustering.

There are many research directions in personal name disambiguation. Song et al. [1] exploited two topic-based models to extract features from corpus and achieved a good effect for personal name disambiguation. Zhao et al. [2] made use of the personal ontology to complete feature extraction and similarity calculation on two real datasets, where the highest similarity is selected for disambiguation. Xu et al. [3] utilized a network embedding-based technique to disambiguate the author name, in which networks are created from papers that have a target ambiguous author name. Yu and Yang [4] solved the challenging task under the circumstances of inadequate data sources. A feature learning means and an affinity propagation clustering were taken into account. Kim et al. [5] combined global features with structure features for author name disambiguation. Global features, extracted from attributes of dataset, formed the textual vector representation. Moreover, negative samples were employed to train a global model. Protasiewicz and Dadas [6] produced a hybrid framework considering both rule-based method and agglomerative hierarchical clustering. Rules were generated from the knowledge of experts, analysis, and so forth. A function  $C_{index}$  was also proposed to determine the best threshold for stopping the hierarchical clustering algorithm. Du et al. [7] applied spectral clustering to recognize ambiguous names in large-scale scientific literature datasets. A distributed approach using Spark framework was advanced to perform in large-scale datasets. Pooja et al. [8] concentrated on the namesake issue of author name disambiguation. They presented an ATGEP method by taking advantage of a graph theory combined with an edge pruning operation.

The research on personal name disambiguation in Chinese datasets is also studied among a number of scientists. Chen et al. [9] provided a feature weighting scheme by calculating pointwise mutual information between personal name and feature word. A trade-off indicator was designed to measure the quality of clusters and stop hierarchical clustering. Li and Wang [10] developed a multistage clustering algorithm based on the entity knowledge, which can be used for Chinese named entity recognition and disambiguation. Ke et al. [11] handled the author name disambiguation under the condition of insufficient information and missing data. Their algorithm devised a novel combination of indicator and incorporated back propagation neural networks.

### 3. Framework and Methodology

*3.1. Research Framework.* The research framework of disambiguating personal name is depicted in Figure 1. There are four main parts in the framework: preprocessing, feature selection, clustering, and labeling. The detailed procedure can be presented in the following steps.

- (1) Preprocess the raw corpus by removing the html tags
- (2) Perform the lexical analysis, including Chinese word segmentation, POS tagging, and named entity recognition
- (3) Select words as features and build feature vectors to represent documents
- (4) Chose the weighting scheme for feature vectors
- (5) Calculate similarity between documents and perform the clustering algorithm
- (6) Evaluate the clustering results and assign each cluster a label

*3.2. Feature Selection.* Extracted features should be capable of distinguishing between people with the same name. The features selected in this paper are outlined as follows:

- (i) Feature 1: named entities (NE)
- (ii) Feature 2: nouns (N)
- (iii) Feature 3: nouns and verbs (N + V)
- (iv) Feature 4: nouns with their document frequency ( $df > 1$ ) (N + Df1)
- (v) Feature 5: named entities with name extension (NE + NameEx)
- (vi) Feature 6: nouns with name extension (N + NameEx)
- (vii) Feature 7: nouns and verbs with name extension (N + V + NameEx)
- (viii) Feature 8: nouns with their  $df > 1$  and name extension (N + Df1 + NameEx)

where  $df$  represents the count of documents having a certain term. A word with  $df = 1$  should be ignored since it cannot contribute to the discrimination between documents.

Due to the weakness of the Chinese word segmentation tool, the personal name in the document may not be correctly identified. A series of rules for name extension are devised according to the results of word segmentation, which is shown in Table 1. Part-of-speech “nr,” “nr1,” “nr2,” “nrf,” and “ng” represent “personal name,” “Chinese surname,” “Chinese given name,” “transcribed personal name,” and “noun morpheme.” “w” denotes “punctuation mark.” The extension of personal name improves the accuracy of target name recognition.

Also, feature selection can be performed in the whole document (Document) or the paragraphs (Paragraph) encompassing the target personal name. Different schemes give birth to different results.

*3.3. Feature Weight.* As documents are represented by a vector space model (VSM), each feature vector can be obtained by calculating the feature weight of a document. A variety of weighting schemes are raised in previous work. This

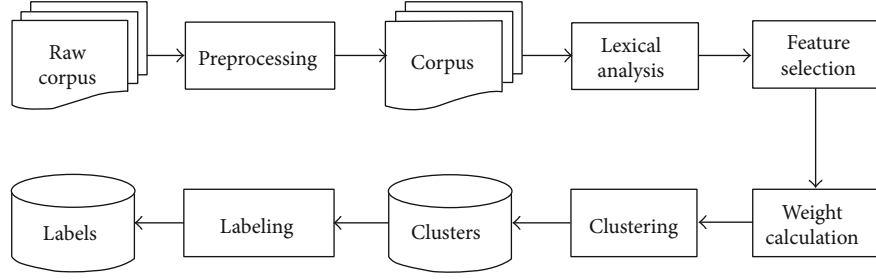


FIGURE 1: Research framework of personal name disambiguation.

TABLE 1: Rules for personal name extension (nr is a Chinese part-of-speech representing personal name).

Rule no.	Pattern	Examples
Rule 1	nr1 + n => nr	Li(李)/nr1 Jun(军)/n => Li Jun(李军)/nr
Rule 2	nr + nr2/ng => nr	Li Jun(李军)/nr Shi(师)/ng => Li Junshi(李军师)/nr
Rule 3	nrf + w + nrf => nr	Roger(罗杰)/nrf./w Musson(穆森)/nrf => Roger Musson(罗杰.穆森)/nr

research adopts four different types of weight calculation methods, which are discussed in detail as follows.

- (i) *Boolean Weights*. A weight assigned to a word is either 0 or 1.0 represents absence of a word in the document, whereas 1 represents the presence. Formula (1) displays the mathematical expression of Boolean weights, where  $f_{ij}$  indicates the frequency of word  $i$  existing in document  $j$ .

$$w_{ij} = \begin{cases} 1 & \text{if } f_{ij} > 0, \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

- (i) *Frequency Weights*. This weighting scheme gives each word an absolute frequency, which is the number of occurrences of word  $i$  in document  $j$ .

$$w_{ij} = f_{ij} \quad (2)$$

- (i) *tf-idf Weights*. A word that appears in only a few documents is likely to be a better discriminator than one that occurs in most or all documents. Inverse document frequency (idf) gives greater weight to words that appear in fewer documents. The tf-idf weight assigned to word  $i$  in document  $j$  can be calculated by formula (3).

$$w_{ij} = tf_{ij} \times idf_i = tf_{ij} \times \log \frac{N}{df_i}, \quad (3)$$

where  $tf_{ij}$  is  $f_{ij}$  divided by total number of words in the document.  $N$  is the count of documents in entire collection and  $df_i$  is the number of documents with word  $i$

- (i) *Entropy Weights*. The entropy weight method introduces the concept of entropy to measure the distribution of words  $i$  in document  $j$ , so the basic idea of the entropy method is similar to idf. It can be defined as follows formula (4):

$$w_{ij} = \log(tf_{ij} + 1) \times \left( 1 + \frac{1}{\log N} \sum_{j=1}^N \frac{tf_{ij}}{n_i} \log \left( \frac{tf_{ij}}{n_i} \right) \right) \quad (4)$$

### 3.4. Clustering Algorithm

**3.4.1. Hierarchical Clustering.** Hierarchical clustering can be divided into two types: divisive and agglomerative. This paper chose the latter because the complexity of divisive clustering algorithm is relatively high and not practical for this task.

Agglomerative hierarchical clustering belongs to a bottom-up method [12]. It treats each document containing target personal name as a separate cluster in the beginning. The algorithm merges two most similar clusters into a larger one at each step until the maximum similarity of clusters exceeds a preset threshold or there is only one cluster left. For the similarity formula, this paper calculates cosine of the angle between the vectors  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  and  $\mathbf{y} = (y_1, y_2, \dots, y_n)$  [13]. It can be written as formula (5).

$$\text{sim}(\mathbf{x}, \mathbf{y}) = \cos(\mathbf{x}, \mathbf{y}) = \frac{\mathbf{x} \cdot \mathbf{y}}{|\mathbf{x}| \cdot |\mathbf{y}|} = \frac{\sum_{i=1}^n x_i \cdot y_i}{\sqrt{\sum_{i=1}^n x_i^2} \cdot \sqrt{\sum_{i=1}^n y_i^2}}. \quad (5)$$

**3.4.2. K-Means Clustering.** K-means algorithm selects  $k$  documents as centroids to form initial clusters. Each document is

TABLE 2: Data statistics for each personal name in corpus.

Personal name	Number of documents	Discarded (gold standard)	Discarded (ICTCLAS)	Discarded (LTP)
Li Jun (李军)	234	1	0	4
Roger (罗杰)	357	24	3	2
Gao Jun (高军)	300	82	84	107
Sun Ming (孙明)	207	2	2	2
Zhang Jianjun (张建军)	247	0	0	1

repeatedly assigned to different clusters according to the closest centroid. Then, the centroid of each cluster will be recomputed. The iteration stops when a convergence criterion is satisfied or after a fixed number of iterations.

**3.4.3. Spectral Clustering.** Spectral clustering is a type of graph-based clustering. It utilizes the eigenvalues or spectrum of the similarity matrix to achieve the goal of dimensionality reduction. Documents can be assigned to different clusters based on the lower-dimensional representation. There are three basic stages in spectral clustering, including preprocessing, decomposition, and grouping.

**3.4.4. GMM Clustering.** Gaussian mixture models (GMM) clustering, also known as expectation-maximization (EM) clustering, makes use of the optimization strategy to cluster unlabeled documents. GMM assumes that data are generated by a Gaussian distribution and tries to obtain a mixture of multidimensional Gaussian probability distributions which can best model any dataset.

**3.5. Labeling Approach.** In order to summarize the person information of each cluster produced by the clustering algorithm, a labeling step is necessary. A simple way of creating a label is to choose a group of representative feature words by ranking the weights of all feature words in cluster.

The labeling algorithm [9] combines mutual information (MI) with  $tf$  to score the weights. For each feature word  $x_i$  in cluster  $C_k$ , the score is calculated by formula (6).  $MI(x_i, \text{name})$  measures the mutual information between the feature word and personal name.  $tf(x_i, C_k)$  counts the number of  $x_i$  appearing in cluster  $C_k$ . We can acquire a label of  $k$  words by taking the top  $k$  feature words in the scoring process.

$$\text{score}(x_i, C_k) = MI(x_i, \text{name}) \times MI_{\text{name}}(x_i, C_k) \times (1 + \log(tf(x_i, C_k))), \quad (6)$$

$$MI(x_i, \text{name}) = \frac{p(x_i, \text{name})}{p(x_i) \times p(\text{name})} = \frac{df(x_i, \text{name}) \times N}{df(x_i) \times df(\text{name})}, \quad (7)$$

$$MI_{\text{name}}(x_i, C_k) = \frac{p(x_i, C_k)}{p(x_i) \times p(C_k)} = \frac{df(x_i, C_k) \times N}{df(x_i) \times df(C_k)}. \quad (8)$$

## 4. Experiment and Discussion

**4.1. Dataset.** The dataset (the dataset is from CLP-2010.) is composed of 1345 files with html tags, including 109 discarded documents that do not have correct target personal

TABLE 3: Comparison of clustering algorithms (feature 1 + paragraph + tf).

Clustering algorithm	Precision	Recall	F score
Hierarchical	78.04%	89.76%	<b>80.94%</b>
K-means	77.44%	77.68%	74.94%
Spectral	68.03%	91.74%	74.16%
GMM	77.34%	82.24%	77.91%

names. The contents of documents are from ‘‘People’s Daily.’’ There are five personal names and each of them contains 200-400 news corpus.

Chinese word segmentation, parts-of-speech tagging, and named entity recognition are performed on corpus. Two types of segmentation and tagging toolkits are exploited: ICTCLAS (<http://ictclas.nlpir.org/>) and LTP(<http://ltp.ai>). As the performance of word segmentation has an important impact on the accuracy of personal name recognition, we compared the number of discarded documents for two toolkits (see Table 2). Gold standard gives the real number of discarded documents. Result shows the personal name recognition of ICTCLAS is more precise than LTP.

**4.2. Evaluation.** Purity and inverse purity [14, 15] are taken as precision and recall for evaluating the clustering effect. Suppose  $S$  is the cluster set to be evaluated and  $R$  is the manually labeled category set. The definition of purity and inverse purity can be described by formula (9) and (10).

$$\text{Precision} = \text{Purity} = \frac{\sum_{S_i \in S} \max_{R_j \in R} |S_i \cap R_j|}{\sum_{S_i \in S} |S_i|}, \quad (9)$$

$$\text{Recall} = \text{InversePurity} = \frac{\sum_{R_i \in R} \max_{S_j \in S} |R_i \cap S_j|}{\sum_{R_i \in R} |R_i|}. \quad (10)$$

The  $F$  score calculates the harmonic mean of precision and recall, which is defined by formula (11). The overall  $F$  score of five personal names is the average of these values.

$$F\text{score} = \frac{2\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (11)$$

## 5. Result and Discussion

The Chinese personal name is recognized by the lexical analysis tool ICTCLAS in this research. We classify all documents containing the same personal name into one

TABLE 4: Comparison of feature selection (tf-idf).

Features	Document			Paragraph		
	Precision	Recall	<i>F</i> score	Precision	Recall	<i>F</i> score
NE	80.15%	92.24%	83.32%	77.74%	88.84%	80.36%
N	81.38%	93.74%	84.59%	82.49%	92.85%	85.15%
N + V	80.10%	93.83%	83.83%	80.93%	93.11%	83.86%
N + Df1	82.70%	92.76%	85.45%	82.77%	90.40%	84.75%
NE + NameEx	89.32%	92.55%	90.87%	86.81%	88.94%	87.79%
N + NameEx	91.07%	94.22%	<b>92.58%</b>	91.29%	93.12%	<b>92.15%</b>
N + V + NameEx	89.57%	94.31%	91.82%	90.55%	93.45%	91.92%
N + Df1 + NameEx	91.09%	93.25%	92.11%	90.28%	90.65%	90.40%

TABLE 5: Comparison of feature weight calculating approach (with highest *F* score).

Feature weights	Document			Paragraph		
	Precision	Recall	<i>F</i> score	Precision	Recall	<i>F</i> score
Bool (N + NameEx)	86.28%	93.99%	89.8%	90.66%	94.59%	<b>92.51%</b>
Tf (N + NameEx)	89.90%	93.44%	91.59%	91.00%	93.39%	92.11%
Tf-idf (N + NameEx)	91.07%	94.22%	<b>92.58%</b>	91.29%	93.12%	92.15%
Entropy (N + Df1 + NameEx)	90.54%	92.49%	91.47%	89.63%	86.88%	88.04%

directory and discard documents that do not contain a target personal name. This paper selects NE features involved in paragraphs containing the target name and frequency weights for testing different clustering approaches. The number of clusters  $k$  is extracted as a prior value from the gold standard. Results of four clustering algorithms are displayed in Table 3.

From Table 3, the hierarchical clustering algorithm outperforms other methods. Additional experiments utilizing other features showed similar results and verified the advantage of hierarchical clustering.

After choosing the clustering algorithm, different combinations of features are adopted for comparison. Table 4 summarizes the clustering results with tf-idf weights.

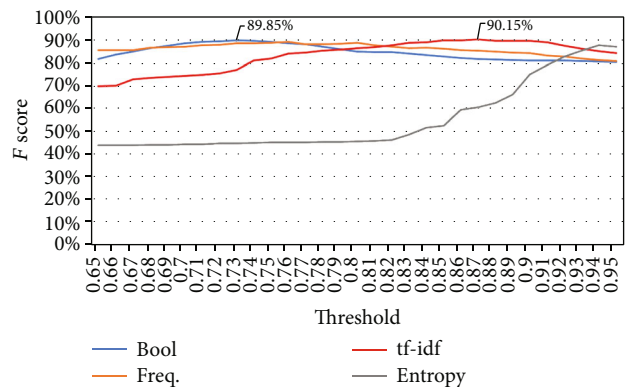
Named entities (NE) are reckoned as the baseline in this experiment. The *F* score increases significantly when other noun features are included (NE belongs to noun). However, adding verb features will lead to a drop in *F* score in comparison with only using noun features. Even though some verbs help to identify the identity of person, most verb features are very limited in disambiguating the personal name. A large number of unrelated verbs will bring noisy data in features, leading to poor experimental results.

Owing to the inaccuracy of word segmentation tools, personal names in documents cannot be effectively identified. When the recognition of an ambiguous name fails, documents that should be clustered are discarded. Therefore, personal name extension is introduced by setting some rules to identify target names. The results presented in Table 4 suggest that name extension dramatically improves the *F* score of clustering results.

The scope of feature selection can be either in the document or in each paragraph having the ambiguous name. According to Table 4, the feature selection in the whole doc-

TABLE 6: Result of personal name disambiguation (N + NameEx +document+tf-idf).

Personal name	Precision	Recall	<i>F</i> score
Li Jun (李军)	84.19%	86.32%	85.24%
Roger (罗杰)	82.20%	92.10%	86.87%
Gao Jun (高军)	100.0%	98.62%	99.31%
Sun Ming (孙明)	94.63%	98.52%	96.54%
Zhang Jianjun (张建军)	94.33%	95.55%	94.94%
Average	91.07%	94.22%	92.58%

FIGURE 2: Distribution of *F* score at different thresholds.

ument yields better results than in paragraph with the exception of noun feature.

Results of four feature weighting schemes are shown in Table 5 with highest *F* score. The corresponding features are listed in parenthesis in the first column. As can be seen



TABLE 7: Labels for “Gao Jun(高军)” clusters.

Entity	Number of documents	Produced labels
Gao Jun 1	208	选手(player), 金牌(gold medal), 冠军(champion), 奥运会(Olympic games), 朝鲜(North Korea), 邓亚萍(Deng Yaping), 队(team), 乔红(Qiao Hong), 女子(women), 乒乓球(table tennis)
Gao Jun 2	4	大队(brigade), 柬埔寨(Cambodia), 金边(Phnom Penh), 工程兵(engineer), 华人(Chinese), 桥梁(bridges), 大队长(captain), 运动会(games), 工兵(engineer), 李金勇(Li Jinyong)
Gao Jun 3	1	编辑(editor), 劳有林(Lao Youlin), 评论(comment), 李东生(Li Dongsheng), 吉菲(Ji Fei), 窦文涛(Dou Wentao), 民委(civil affairs committee), 大桥(bridge), 电视(TV), 杜晓春(Du Xiaochun)
Gao Jun 4	2	王圣珍(Wang Shengzhen), 选集(anthology), 深情(affection), 毛选(Mao Xuan), 字字句句(words and sentences), 写字台(writing desk), 手抄本(manuscripts), 春秋(spring and autumn), 日日夜夜(days and nights), 书法(calligraphy)

from the table, results obtained by tf weight in the whole document are better than others.

The detail result of every personal name is described in Table 6 when the average  $F$  score for all names achieves a best value.

Clustering algorithm stops when the similarity between two clusters is less than a certain threshold. The relationship between threshold and  $F$  score can be illustrated in Figure 2. The influence of threshold on results depends on feature sets. We choose feature 6 (N + NameEx) plus document to run clustering. The value of threshold is given through enumeration by every 0.01 step. The  $F$  score reaches a highest value of 90.15% when tf-idf weight is selected.

Basic information about a person is given by labeling process. For instance, clusters of Gao Jun(高军) are labeled with meaningful words in Table 7. The created labels are representative words that can summarize the characteristics of a person.

## 6. Conclusions

This paper studied the task of Chinese personal name disambiguation based on an unsupervised method. The open dataset contains five ambiguous names with gold standard. We exploited lexical analysis toolkits to perform segmentation and POS tagging. Eight groups of features are selected to combined with four feature weight calculating methods. In order to refine the precision of personal name recognition, name extension is proposed. The extension process of personal name significantly enhances the final effect of clustering experiments. Besides, the agglomerative hierarchical clustering algorithm is chosen from four methods for disambiguating names. The threshold of hierarchical clustering is also tested for different feature weights. At last, labels are constructed for clusters of target name by scoring the weights of feature words in clusters.

Final experimental results demonstrated the effectiveness of the proposed research approach. Nonetheless, some disadvantages may exist in the framework. Rules of personal name extension are suited for the current dataset. It may be necessary to add extra rules for other corpus so as to increase the precision of detecting Chinese personal names. In addition, we will develop automatic feature selection algorithms as well as new weigh calculating methods in future work. More

sophisticated clustering and supervised document classification methods will also be taken into consideration.

## Data Availability

The original dataset used in this work is available from the corresponding author on request.

## Conflicts of Interest

The authors declare no conflicts of interest.

## Acknowledgments

This work was supported by the Youth Foundation of Basic Science Research Program of Jiangnan University, 2019 (No. JUSRP11962) and High-level Innovation and Entrepreneurship Talents Introduction Program of Jiangsu Province of China, 2019.

## References

- [1] Y. Song, J. Huang, I. G. Councill, J. Li, and C. L. Giles, “Efficient topic-based unsupervised name disambiguation,” in *Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL)*, pp. 342–351, Vancouver, BC, Canada, 2007.
- [2] Z. Lu, Z. Yan, and L. He, “Ontology-based personal name disambiguation on the web,” in *2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, pp. 185–192, Atlanta, GA, USA, 2013.
- [3] J. Xu, S. Shen, D. Li, and Y. Fu, “A network-embedding based method for author disambiguation,” in *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pp. 1735–1738, Torino, Italy, 2018.
- [4] Z. Yu and B. Yang, “Researcher name disambiguation: feature learning and affinity propagation clustering,” in *International Symposium on Methodologies for Intelligent Systems*, pp. 225–235, Springer, 2018.
- [5] K. Kim, S. Rohatgi, and C. L. Giles, “Hybrid deep pairwise classification for author name disambiguation,” in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, pp. 2369–2372, Beijing, China, 2019.
- [6] J. Protasiewicz and S. Dadas, “A hybrid knowledge-based framework for author name disambiguation,” in *2016 IEEE*

- International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 000594–000600, Budapest, Hungary, 2016.
- [7] H. Du, Z. Jiang, and J. Gao, “Who is who: name disambiguation in large-scale scientific literature,” in *2019 International Conference on Data Mining Workshops (ICDMW)*, pp. 1037–1044, Beijing, China, 2019.
  - [8] K. M. Pooja, S. Mondal, and J. Chandra, “A graph combination with edge pruning-based approach for author name disambiguation,” *Journal of the Association for Information Science and Technology*, vol. 71, no. 1, pp. 69–83, 2020.
  - [9] C. Chen, H. Junfeng, and W. Houfeng, “Clustering technique in multi-document personal name disambiguation,” in *Proceedings of the ACL-IJCNLP 2009 Student Research Workshop*, pp. 88–95, Suntec, Singapore, 2009.
  - [10] G. Li and H. Wang, “Chinese named entity recognition and disambiguation based on multi-stage clustering,” *Journal of Chinese Information Processing*, vol. 27, no. 5, pp. 29–34, 2013.
  - [11] H. Ke, T. Li, Y. Zhou, Y. Zhong, Z. Yu, and J. Yuan, “Aauthor name disambiguation using BP neural networks under missing data,” *Journal of the China Society for Scientific and Technical Information*, vol. 37, no. 6, pp. 600–609, 2018.
  - [12] C. Schätzle and H. Booth, “DiaHClust: an iterative hierarchical clustering approach for identifying stages in language change,” in *2019 in Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change*, pp. 126–135, Florence, Italy, 2019.
  - [13] Q. Zhou and L. Leydesdorff, “The normalization of occurrence and co-occurrence matrices in bibliometrics using cosine similarities and Ochiai coefficients,” *Journal of the Association for Information Science and Technology*, vol. 67, no. 11, pp. 2805–2814, 2016.
  - [14] C. Fan and J. Yu, “Finding community structure in social network of Renren,” *ICIC Express Letters*, vol. 7, no. 5, pp. 1693–1698, 2013.
  - [15] A. Hotho, S. Staab, and G. Stumme, “WordNet improves text document clustering,” in *Proceedings of the SIGIR 2003 Semantic Web Workshop*, pp. 541–544, Toronto, Canada, 2003.

## Research Article

# An Efficient Q-Algorithm for RFID Tag Anticollision

Lingyun Zhao,<sup>1</sup> Lukun Wang ,<sup>1</sup> and Shan Du<sup>2</sup>

<sup>1</sup>College of Intelligent Equipment, Shandong University of Science and Technology, Taian 271019, China

<sup>2</sup>Department of Computer Science, Mathematics, Physics and Statistics, University of British Columbia, Okanagan, V1V 1V7, Canada

Correspondence should be addressed to Lukun Wang; wanglukun@sdust.edu.cn

Received 18 March 2021; Revised 14 April 2021; Accepted 19 April 2021; Published 10 May 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Lingyun Zhao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In large-scale Internet of Things (IoT) applications, tags are attached to items, and users use a radiofrequency identification (RFID) reader to quickly identify tags and obtain the corresponding item information. Since multiple tags share the same channel to communicate with the reader, when they respond simultaneously, tag collision will occur, and the reader cannot successfully obtain the information from the tag. To cope with the tag collision problem, ultrahigh frequency (UHF) RFID standard EPC G1 Gen2 specifies an anticollision protocol to identify a large number of RFID tags in an efficient way. The Q-algorithm has attracted much more attention as the efficiency of an EPC C1 Gen2-based RFID system can be significantly improved by only a slight adjustment to the algorithm. In this paper, we propose a novel Q-algorithm for RFID tag identification, namely, HTEQ, which optimizes the time efficiency of an EPC C1 Gen2-based RFID system to the utmost limit. Extensive simulations verify that our proposed HTEQ is exceptionally expeditious compared to other algorithms, which promises it to be competitive in large-scale IoT environments.

## 1. Introduction

RFID is a key enabler of the Internet of Things (IoT), playing a crucial role in connecting low-/nonpowered devices to IoT environments. EPC C1 Gen2 [1] is the standard RFID protocol devised to meet the needs of such applications. An RFID reader can communicate with hundreds of passive (nonpowered) tags within seconds, even at a distance of 10 meters from the tags. The most remarkable virtue of the Gen2 standard is its light weight. Due to the shared nature of RF medium, a passive RFID system requires a collision arbitration protocol to serialize responses of tags, mitigating collisions between the tag responses.

Recently, there are mainly three kinds of collision arbitration protocols in tag identification, namely, Aloha-based [2–5], query tree-based [6–8], and tree splitting-based [9, 10] protocol. In the Aloha-based protocol, the reader combines tag number estimation and frame size adjustment strategies to identify tags. Specifically, the reader sends a query command containing a parameter  $F$  ( $F$  denotes the number of) to allow the tags to return the IDs. The reader detects the

responses in each slot and distinguishes their different states: collision, empty, or singleton. According to slot statistics in a frame, the reader can estimate the cardinality of unread tags and update the new frame for the next round. Through theoretical analysis, the maximal system throughput of Aloha-based protocols is 0.368 [2–5]. In query tree-based protocols [6–8], the reader inquires about the tags through probe commands. Each tag is required to be equipped with a prefix matching circuit, and it will respond only when the tag ID matches the prefix of the probe command. Once a collision is detected, the reader will update the query prefix according to the position of collision bits. Then, the reader uses the updated prefixes to interrogate the tags until all of them are successfully identified. In tree splitting protocols, the reader continues to group the colliding tags with a separation probability of 0.5, until a certain group contains only one tag.

The Q-algorithm, adopted by the EPC C1 Gen2 standard, is a collision arbitration algorithm that clearly represents its lightweight property such that a couple of simple arithmetic operations constitute the algorithm. The Q-algorithm is composed of a few fundamental arithmetic and logical

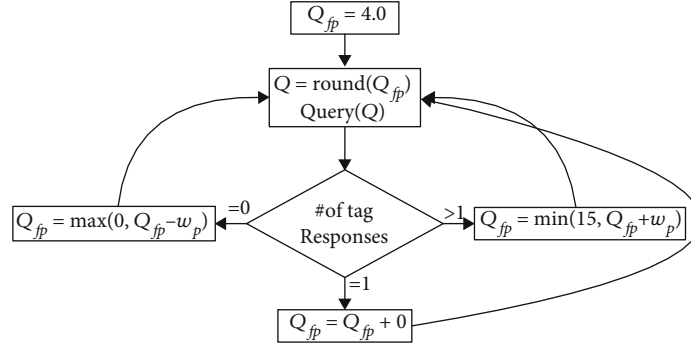


FIGURE 1: The flowchart of the Q-algorithm.

operations [1]. The flowchart of the Q-algorithm is illustrated in Figure 1. Looking specifically into the algorithm, it adds the constant  $c$  to  $Q$  when it finds collided replies and subtracts the  $c$  constant from  $Q$  whenever having no reply. The algorithm alters its frame size  $2^Q$  once the  $Q$  value has been changed. This simplicity of the logic and the use of a minimal number of computations make Gen2 RFID readers identify tags without wasting batteries in IoT environments. Another strength inherited in the Q-algorithm is the early adjustment of frame size, by which a disadvantageous frame size can be canceled before the end of the frame. The feature might have a substantial impact on the identification performance in large-scale and/or mobile tag environments such as smart factories and highways, which is an important characteristic of a quality collision arbitration algorithm [11].

Only a limited number of early researchers attempted to further improve the performance of the Q-algorithm [12–14]. The  $Q^+$  algorithm [12] suggested differentiating  $c$  into  $C_{\text{coll}}$  and  $C_{\text{idle}}$ , which means the constant for a collision and an empty slot, respectively. The study also found the optimal ratio of  $C_{\text{idle}}$  to  $C_{\text{coll}}$  ( $e^{-2} \approx 0.71828$ ) which maximizes the per-slot throughput of tag identification. In a similar vein, the author in literature [13] presents a fine-tuned numerical method to determine the  $Q$  value. However, the study also considered a per-slot throughput as its objective, which means that, theoretically, the performance of such an algorithm is not very different from that of the  $Q^+$  algorithm.

Another approach has been taken to improve the Q-algorithm in [14], where a per-time throughput is addressed as the metric of tag identification efficiency. This is important as the time duration of each slot (singleton, empty, and collision) is actually different from each other in the EPC C1 Gen2 standard. As the studies commonly recognized the difference, the algorithms were designed to determine  $C_{\text{idle}}$  and  $C_{\text{coll}}$  that reduce the collision slots, whose time duration is significantly longer than the empty slots, rather than the total number of consumed slots. Although such an algorithm is a reasonable approach, the rigorous optimization of the Q-algorithm in terms of time efficiency has not been addressed. Due to such immaturity of the previous research efforts, the Q-algorithm has received little attention for a good while, despite its obvious benefits of lightweight property and innate ability of early adjustment.

On the other hand, there are many works following the dynamic framed slotted Aloha (DFSA) to consider the time efficiency. The literature [15] conducted temporal analysis on the time efficiency of an FSA protocol, adopting an in-frame adjustment feature of the Q-algorithm. The performance of conventional DFSA depends on both an accurate cardinality estimation and adaptive frame adjustment. The cardinality estimation requires burdensome statistics and imposes high complexity. The latest research [3, 16] shows remarkable performance with optimization of frame sizes for fast identification; additional calculations for tag cardinality estimation were required on every slot.

In this paper, we provide a theoretical analysis on the time performance of the Q-algorithm, and based on the analysis, the detailed structure of a highly time-efficient Q-algorithm (HTEQ) is proposed. Then, we provide a performance evaluation between major Q-algorithm variants and a competitive DFSA algorithm. In the following, we further discuss the experimental results.

The contributions of this paper are summarized as follows.

- (1) A rigorous analysis of the time performance of the Q-algorithm and the optimal parameters for a time-efficient Q-algorithm are derived
- (2) We provide a thorough performance evaluation between the proposed HTEQ, other competitive Q-algorithm variants, and representative DFSA. The factors behind the experimental results are examined concretely
- (3) Through performance limitation analysis, a novel RFID tag identification algorithm, namely, HTEQ, is proposed, which provides highly competitive time efficiency in RFID anticollision for large-scale IoT applications, while maintaining the strengths of the Q-algorithm, such as early frame adjustment and low computational overhead

## 2. The Proposed Algorithm

In this section, we first analyze the time performance of an FSA protocol on which the Q-algorithm is premised and

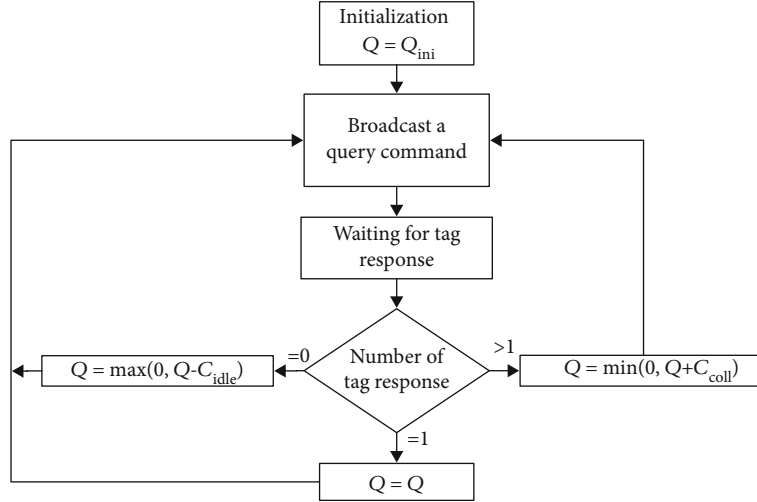


FIGURE 2: The flowchart of the proposed HTEQ algorithm.

suggest a highly efficient  $Q$ -algorithm based on theoretical analysis.

**2.1. Analysis on Time Performance.** In the FSA protocol, a reader informs tags of a frame size, in which the tags randomly select their responding slots. In such a model, the tag identification delay can be defined as the average time between two successively identified tags. Given  $n$  tags, the tag identification delay in a frame size  $F$  is expressed as

$$\eta_{n,F} = \frac{T_{n,F}}{F \times P_s}, \quad (1)$$

where  $T_{n,F}$  and  $P_s$  denote the expected time duration of a frame and the probability that a tag is successfully identified in a slot given  $n$  and  $F$ , respectively. The probability that  $k$  tags respond in a slot is given by

$$P(x=k) = C_n^k \left(\frac{1}{F}\right)^k \left(1 - \frac{1}{F}\right)^{n-k}. \quad (2)$$

Based on Equation (2), the probabilities of a success, an empty, and a collision slot are expressed as  $P_s = P(x=1)$ ,  $P_e = P(x=0)$ , and  $P_c = P(x>1)$ , respectively. Then, the expected time duration of a frame can be calculated by summing up all the expected delays in the frame as below:

$$T_{n,F} = F \times P_s \times T_s + F \times P_e \times T_e + F \times P_c \times T_c, \quad (3)$$

where  $T_s$ ,  $T_e$ , and  $T_c$  are the time durations of a success, an empty, and a collision slot, respectively. Note that the delay of the frame per second is ignored for convenience's sake.

TABLE 1: Parameters used for various algorithms.

Algorithms	$C_{coll}$	$C_{idle}$	$C_{idle}/C_{coll}$
HTEQ	0.2118	0.0827	0.3906
fastQ	0.2118	0.1500	0.7081
$Q^+$ algorithm	0.2118	0.1522	0.7182
$Q$ -algorithm	0.2118	0.2118	1

By Equations (2) and (3), Equation (1) can be rewritten as

$$\begin{aligned} \eta_{n,F} &= \frac{F \times P_s \times T_s + F \times P_e \times T_e + F \times P_c \times T_c}{F \times P_s} \\ &= T_s + \left( \frac{F-1}{n} (\lambda - 1) + \frac{F}{n} \left(1 - \frac{1}{F}\right)^{1-n} - 1 \right) \times T_c, \end{aligned} \quad (4)$$

where  $\lambda = T_e/T_c$ .

**2.2. HTEQ Algorithm.** Having observed the tag identification delay, we can achieve the maximization of per-time tag identification efficiency by minimizing the tag identification delay as the following:

$$\arg \min_F \eta_{n,F}. \quad (5)$$

As the objective function is convex for both  $n$  and  $F$ , the optimum should be obtained by partially differentiating the function for  $n$  ( $n \geq 1$ ) instead of  $F$  on which the inverse of the derivative does not exist. Therefore, arranging  $\partial\eta/\partial\eta = 0$  with respect to  $F$ , we get the optimal frame size as below:

$$F^* = \left(1 - e^{-1+W(\Theta)/n}\right)^{-1}, \quad (6)$$

where  $W()$  is the Lambert  $W$  function or product logarithm and  $\Theta = (\eta - 1)/e$ . We then get  $P_e^*$  and  $P_c^*$  under the optimal condition by putting  $F^*$  to the probabilities. Using these probabilities, we can build a stationary condition that

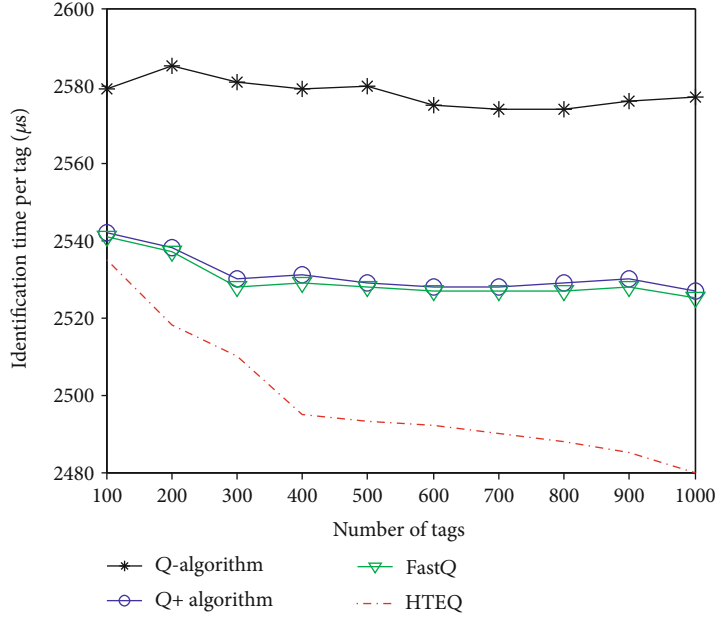


FIGURE 3: Comparison of identify delays of various algorithms.

converges to and keeps the frame size remaining at the optimum as follows:

$$P_c^* \times C_{\text{coll}} = P_e^* \times C_{\text{idle}}. \quad (7)$$

With this relationship, taking limits in  $P_c^*/P_e^*$  as  $n \rightarrow \infty$  for  $0 < \eta < 1$ , it gives us the asymptotic proportions of the empty and collision slots under the optimal condition, and hence, the optimal ratio  $C_{\text{idle}}^*/C_{\text{coll}}^*$  is also attained:

$$\lim_{n \rightarrow \infty} \frac{P_{\text{coll}}^*}{P_{\text{idle}}^*} = \frac{C_{\text{idle}}^*}{C_{\text{coll}}^*} = e^{1+W(\Theta)} - W(\Theta) - 2. \quad (8)$$

Note that the ratio of the probabilities and the ratio of the  $C$  values are in a reciprocal relationship. Figure 2 provides a detailed procedure of the HTEQ algorithm, where  $C_{\text{idle}}$  and  $C_{\text{coll}}$  are constrained by Equation (8). The overall workflow of the proposed algorithm is similar to the Q-algorithm, adhering to the principle of not requiring any tag estimation scheme. The only difference with the Q-algorithm is the biased constants  $C_{\text{idle}}$  and  $C_{\text{coll}}$ .

### 3. Experimental Study

The performance of the proposed approach and the reference methods were examined by carrying out extensive simulations based on the Monte Carlo simulations. The performance of the tag identification algorithms is examined in two evaluation sessions with different comparison groups: (1) Group I for comparison between Q-algorithms, where the original Q-algorithm [1] and its variants in the strict sense of the term, Q<sup>+</sup> [12] algorithm and fast Q [14] algorithm, are selected for the comparison, to prove the potential of the HTEQ algorithm, which relies on tag estimation and

frame size determination strategies; (2) Group II for comparison with DFSA algorithms, where relatively recent, competitive DFSA algorithms, EACAEA [5] and TES-FAS [3], are chosen as the comparison targets. For the protocol parameters in the simulations, we adopt the same setting used in [14]: 1828.13  $\mu\text{s}$ , 260.625  $\mu\text{s}$ , and 516.625  $\mu\text{s}$  are used for  $T_s$ ,  $T_e$ , and  $T_c$ , respectively.

For the fairness of the evaluation, the  $C_{\text{idle}}$  and  $C_{\text{coll}}$  values are recalculated accordingly with the aforementioned protocol parameters. Specifically, the  $C_{\text{idle}}$  and  $C_{\text{coll}}$  values of fastQ remained the same as in [14]. For the sake of clarity of comparison, the  $C_{\text{coll}}$  values of HTEQ and Q<sup>+</sup> algorithm are set to be identical to the values of the fastQ algorithm, and the  $C_{\text{idle}}$  values are calculated using the ratios  $C_{\text{idle}}/C_{\text{coll}}$  suggested by each algorithm. Table 1 summarizes the adjusted parameters. All the algorithms are implemented on MATLAB R2012b, and simulations are iterated up to 2000 times with varying random seeds [9, 17–19].

The simulation results in Figure 3 show the identification delays for Q-algorithm-based collision arbitration algorithms to identify one tag. The simulations are carried out every 100 tags between 100 and 1000 tags to identify. Each result is recorded after 5000 iterations for the convergence of each simulation. In Figure 3, our proposed algorithm HTEQ clearly outperforms other reference algorithms. One interesting fact is that the performance of fastQ lies fairly behind that of HTEQ despite its own protocol parameter being used for the simulation. Most algorithms except the Q-algorithm are designed to be more sensitive to collision slots to quickly escape from undersized frames, which can make a huge overhead. They put more weight on the  $C_{\text{coll}}$  value than  $C_{\text{idle}}$ .

Hence,  $C_{\text{idle}}/C_{\text{coll}} < 1$  holds in various algorithms. This is a rational approach; nevertheless, those algorithms except HTEQ did not consider a strict optimization of the  $C_{\text{idle}}/$

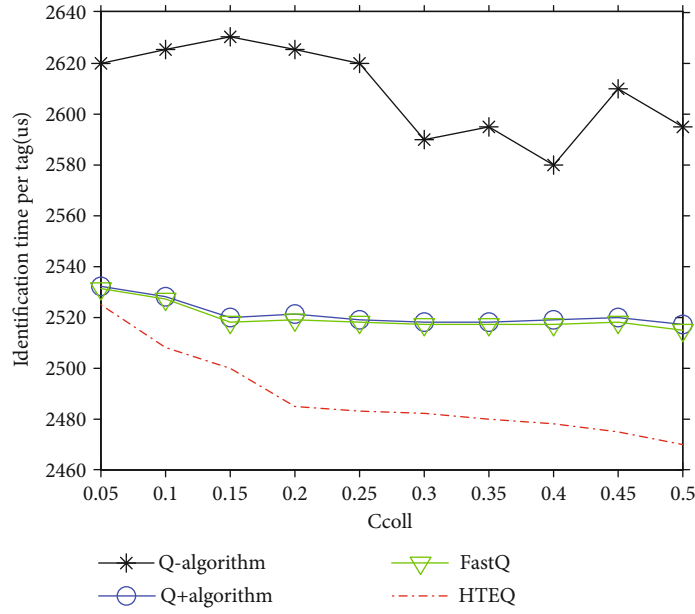


FIGURE 4: Comparison of tag identification time per tag.

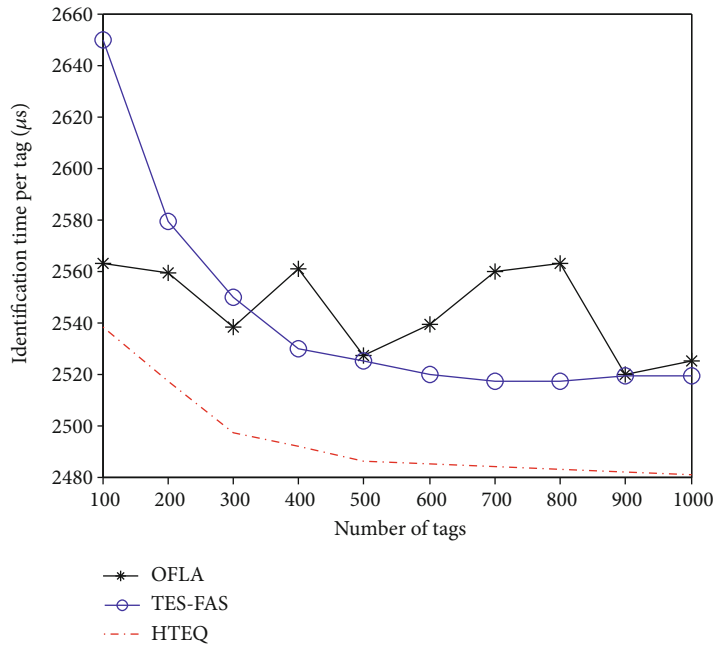


FIGURE 5: Performance comparison to DFSA algorithms.

$C_{coll}$  value. For example, fastQ adopts 0.7081 for its  $C_{idle}/C_{coll}$  value, but the optimal ratio is almost half of this value (0.3906), i.e., the fastQ algorithm may be fast, but not the fastest.

The findings from an exhaustive simulation shown in Figure 4 substantiate the contention above. The simulation results form an obvious convex shape on the  $C_{coll}-C_{idle}/C_{coll}$  plane. The HTEQ algorithm is situated around the optimum area of the convex. The fastQ lies on the higher position  $C_{idle}/C_{coll} = 0.7081$  and followed by the  $Q^+$  algorithm closely. This explains why fastQ and  $Q^+$  algorithm showed similar performance in Figure 3. As a result, it validates that the pro-

posed HTEQ algorithm essentially achieved the highly efficient tag identification with the optimized parameter  $C_{idle}/C_{coll}$ .

As for the simulation results from Group II, shown in Figure 5, HTEQ performs higher, where it consistently outdistances the other algorithms along with all the tag numbers. Given that OFLA and TES-FAS are exquisite algorithms designed to pursue the optimal frame size in regard to time efficiency, the difference in the performance is rather extraordinary. According to our in-depth analysis, both DFSA algorithms suffered from degradation due to the inaccurate tag cardinality estimation. As for OFLA, it is affected by

imperfect frame size calculation in addition to the erroneous tag cardinality estimation.

OFLA followed a similar process to HTEQ for finding the optimal frame size; however, it ended up with the use of the following approximation derived from a numerical method:

$$F_{\text{opt}} = a \times m, \quad (9)$$

where  $a$  is a linear approximation coefficient and  $a = 1.44$  is set for this environment. This is the reason that the performance of OFLA fluctuated along with the tag numbers. The time efficiency of OFLA rose and fall repeatedly according to whether Equation (9) gets closer to or away from the optimal curve which is a nonlinear curve as presented in Equation (6). In the case of TES-FAS, the degradation comes mostly from its inaccurate tag cardinality estimation. The maximum a posteriori estimator used in TES-FAS is known for a highly accurate tag cardinality; however, the subframe-based scheme and its own early frame adjustment feature caused the performance deterioration. The subframe-based scheme reassesses a current frame size whenever reaching its subframe sizes. As the small frame sizes given by the subframe-based scheme reduce sample sizes for tag cardinality estimation, significant estimation errors are accumulated. Although early frame adjustment may give an edge to TES-FAS over the evaluation, especially in the large-scale environment given in this paper, it turned out that the algorithm is caught between the trade-offs due to the degradation of tag cardinality estimation.

On the other hand, HTEQ successfully found a closed-form solution for the optimal frame size and suggested the optimal relationship to determine  $C_{\text{idle}}$  and  $C_{\text{coll}}$  while maintaining the merits of the Q-algorithm. Note that this result is achieved without the consideration of computational cost, in which HTEQ is the most advantageous over the other DFSA algorithms. We assumed that the computational cost is neglectable in conceding that the excessive computational cost of the DFSA algorithms such as the MAP estimator can be reduced to that of the Q-algorithm-based algorithms by using a lookup table and additional memory space (0.3674  $\mu\text{s}$  and 0.8451  $\mu\text{s}$ , respectively, based on the computing power and computational overhead provided in [3]). The computation time that is amounting to 828  $\mu\text{s}$  should be added to the identification time per tag of TES-FAS unless the lookup table is considered. In other words, HTEQ achieved the true optimal performance with a considerably lower computational cost of the Q-algorithm, without any assistance of additional resources.

#### 4. Conclusion

This paper proposed a novel, highly time-efficient Q-algorithm for collision arbitration in large-scale IoT environments. The time performance of an FSA protocol deployed with the Q-algorithm is investigated, and the optimal parameters to maximize the per-time throughput of the Q-algorithm are derived based on the investigation. Our intensive simulations, including the exhaustive one and the comparison to the state-of-the-art DFSA algorithms, proved that

the proposed algorithm is evidently the most time-efficient algorithm, taking full advantage of the innate strengths of the Q-algorithm.

#### Data Availability

To be frank, I derived the writing material from different journals as provided in the references. A MATLAB tool has been utilized to simulate our concept.

#### Conflicts of Interest

The authors declare that they have no conflicts of interest.

#### References

- [1] EPC, *Radio-Frequency Identity Protocols Generation-2 UHF RFID Standard, Release 2.1*, GS1, Brussels, Belgium, 2018.
- [2] J. Su, R. Xu, S. Yu, B. Wang, and J. Wang, "Idle slots skipped mechanism based tag identification algorithm with enhanced collision detection," *KSII Transactions on Internet and Information Systems*, vol. 14, no. 5, pp. 2294–2309, 2020.
- [3] J. Su, Z. Sheng, A. X. Liu, Z. Fu, and Y. Chen, "A time and energy saving based frame adjustment strategy (TES-FAS) tag identification algorithm for UHF RFID systems," *IEEE Transactions on Wireless Communications*, vol. 19, no. 5, pp. 2974–2986, 2020.
- [4] H. Chen, K. Liu, C. Ma, Y. Han, and J. Su, "A novel time-aware frame adjustment strategy for RFID anti-collision," *Computers, Materials & Continua*, vol. 57, no. 2, pp. 195–204, 2018.
- [5] W.-T. Chen, "Optimal frame length analysis and an efficient anti-collision algorithm with early adjustment of frame length for RFID systems," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 5, pp. 3342–3348, 2016.
- [6] X. Jia, Q. Feng, and L. Yu, "Stability analysis of an efficient anticollision protocol for RFID tag identification," *IEEE Transactions on Communications*, vol. 60, no. 8, pp. 2285–2294, 2012.
- [7] J. Su, Y. Chen, Z. Sheng, Z. Huang, and A. X. Liu, "From M-ary query to bit query: a new strategy for efficient large-scale RFID identification," *IEEE Transactions on Communications*, vol. 68, no. 4, pp. 2381–2393, 2020.
- [8] H. Landaluce, A. Perallos, E. Onieva, L. Arjona, and L. Bengtsson, "An energy and identification time decreasing procedure for memoryless RFID tag anticollision protocols," *IEEE Transactions on Wireless Communications*, vol. 15, no. 6, pp. 4234–4247, 2016.
- [9] J. Su, Z. Sheng, L. Xie, G. Li, and A. X. Liu, "Fast splitting based tag identification algorithm for anti-collision in UHF RFID system," *IEEE Transactions on Communications*, vol. 67, no. 3, pp. 2527–2538, 2019.
- [10] J. Su, Z. Sheng, A. X. Liu, Y. Han, and Y. Chen, "A group-based binary splitting algorithm for UHF RFID anti-collision systems," *IEEE Transactions on Communications*, vol. 68, no. 2, pp. 998–1012, 2020.
- [11] J. Myung, W. Lee, J. Srivastava, and T. K. Shih, "Tag-splitting: adaptive collision arbitration protocols for RFID tag identification," *IEEE Transactions on Parallel and Distributed Systems*, vol. 18, no. 6, pp. 763–775, 2007.
- [12] D. Lee, K. Kim, and W. Lee, "Q<sup>+</sup>-algorithm: an enhanced RFID tag collision arbitration algorithm," in *Proceedings of*



*4th International Conference on Ubiquitous Intelligence Computing (UIC)*, Hong Kong, China, 2007.

- [13] W.-T. Chen, "A feasible and easy-to-implement anticollision algorithm for the EPCglobal UHF class-1 generation-2 RFID protocol," *IEEE Transactions on Automation Science and Engineering*, vol. 11, no. 2, pp. 485–491, 2014.
- [14] J. Teng, X. Xuan, and Y. Bai, "A fast Q algorithm based on EPC generation-2 RFID protocol," in *Proceedings of 6th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM)*, pp. 23–25, Chengdu, China, September, 2010.
- [15] T. F. La Porta, G. Maselli, and C. Petrioli, "Anticollision protocols for single-reader RFID systems: temporal analysis and optimization," *IEEE Transactions on Mobile Computing*, vol. 10, no. 2, pp. 267–279, 2011.
- [16] J. Su, X. Zhao, D. Hong, Z. Luo, and H. Chen, "Q-value fine-grained adjustment based RFID anti-collision algorithm," *IEICE Transactions on Communications*, vol. E99.B, no. 7, pp. 1593–1598, 2016.
- [17] J. Su, R. Xu, S. Yu, B. Wang, and J. Wang, "Redundant rule detection for software-defined networking," *KSII Transactions on Internet and Information Systems*, vol. 14, no. 6, pp. 2735–2751, 2020.
- [18] Z. Zhou, Q. M. J. Wu, Y. Yang, and X. Sun, "Region-level visual consistency verification for large-scale partial-duplicate image search," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 16, no. 2, pp. 1–25, 2020.
- [19] Z. Zhou, Y. Mu, and Q. M. J. Wu, "Coverless image steganography using partial-duplicate image retrieval," *Soft Computing*, vol. 23, no. 13, pp. 4927–4938, 2019.

## Research Article

# User Value Identification Based on Improved RFM Model and $K$ -Means++ Algorithm for Complex Data Analysis

Jun Wu,<sup>1,2</sup> Li Shi,<sup>1</sup> Liping Yang,<sup>2</sup> Xiaxia Niu ,<sup>2</sup> Yuanyuan Li,<sup>2</sup> Xiaodong Cui,<sup>3</sup> Sang-Bing Tsai ,<sup>4</sup> and Yunbo Zhang<sup>5</sup>

<sup>1</sup>College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China

<sup>2</sup>School of Economic and Management, Beijing University of Chemical Technology, Beijing 100029, China

<sup>3</sup>Industrial and Commercial Bank of China Limited, Beijing 100088, China

<sup>4</sup>Regional Green Economy Development Research Center, School of Business, Wuyi University, Wuyishan 354300, China

<sup>5</sup>Datang Carera (Beijing) Investment Co. Ltd., Beijing 100191, China

Correspondence should be addressed to Xiaxia Niu; 15136212624@163.com and Sang-Bing Tsai; sangbing@hotmail.com

Received 8 March 2021; Revised 6 April 2021; Accepted 16 April 2021; Published 3 May 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Jun Wu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In recent years, with the development of machine learning and big data technology, user data has become an important element in the production process of enterprises. For today's e-commerce platforms, the deep mining of user's purchase behavior is helpful to understand user's purchase preferences and accurately recommend products that meet user expectations, which can not only improve user satisfaction but also reduce platform marketing cost. To accurately identify the user value of online purchasing on an e-commerce platform, this paper uses an improved RFM model to extract user features and uses the  $K$ -means++ clustering algorithm to realize user classification. The indicators of the traditional RFM model characterize user features from three angles: recent purchase time ( $R$ ), purchase frequency ( $F$ ), and total consumption amount ( $M$ ). The user group and scenarios studied in this paper are different from the previous literature: (1) the user group is relatively fixed, (2) the consumer goods are relatively single, and (3) the characteristics of repeated purchase are obvious. Therefore, based on the existing literature, this paper extracts the user characteristics studied and improves and models the traditional indicators. Based on the real purchasing data from September to December 2018, it calculates the indicators that improved RFM, empowers the weight to indicators, and finally classifies the value of users by using the  $K$ -means++ algorithm. The experimental results show that the user classification based on the improved RFM model is more accurate than the user classification based on the traditional RFM model, and the improved RFM model can identify the user value more accurately, which provides a strong support for the e-commerce platform to realize the accurate marketing strategy based on big data.

## 1. Introduction

Big data technology is the product of the development of information processing technology. Today is an era of data explosion; the enterprise operation process will produce massive data. Today's big data technology is widely used in various scenarios of enterprise operations, internet of things (IOT) environment [1–6], etc., providing decision support for various decisions of enterprises. Compared with the traditional marketing methods such as questionnaire surveys, the purchase behavior of users often reflects the user's psychological preferences. In today's business competition, cus-

tomers are the main focus of the company to maintain excellent performance. It is found that the cost of acquiring new customers is much more expensive than retaining existing customers. Thus, what companies care most is how to sell more products to existing customers. Nowadays, the number of e-commerce platforms is increasing rapidly, the operating cost is getting higher and higher, and the marketing expenditure is also expanding accordingly. Using the purchase records of users on the platform to understand the decisions made by users in the real environment has become an urgent problem to be solved in the efficient operation of enterprises. At the same time, as a new business model, a community

e-commerce platform has less research on the classification of community e-commerce customers. Based on the above background, this paper provides an effective method for the e-commerce platform to identify customer value and realize a precision marketing strategy.

The research object of this paper is the T-app community e-commerce platform customers; this platform mainly sells cooked food, pasta, and other goods, a total of 134 kinds. The platform focuses on the community traffic, aiming at the needs of community residents, and its customers' purchase behavior has the characteristics of a low single consumption amount and high consumption frequency. The RFM model has good representativeness in reflecting customer value and customer purchase preference and is widely used in financial industry [7], retail [8], insurance and telecommunication [9], education industry [10], and e-commerce industry [11]. The RFM model can well construct the outline of the research object in this paper.

However, in the original RFM model, there is a large randomness in the  $R$  index of the latest consumption time. New customers who have just come into contact with the platform and loyal customers of the platform may have the same performance in the  $R$  index of the latest consumption time, and the model cannot describe the dependence of customers on a single commodity. In order to help the platform describe the multidimensional attributes of customers and more accurate customer value, an improved RFM model with five indicators is proposed.

Therefore, based on the improved RFM model and  $K$ -means++ algorithm, this paper proposes a method suitable for T-app community e-commerce platform user classification. According to the real purchase records of users on the T-app community e-commerce platform, the improved RFM model is used to analyze the data of user purchase records. The  $K$ -means++ clustering algorithm is used to classify and calculate the users of the e-commerce platform. Finally, the classification results are explained, and the corresponding customer value analysis is given combined with the specific indicators of each subdivided user. This study uses a quantitative analysis method to segment and cluster platform customers. Customer segmentation with clear value and purchase preference is helpful for the platform to effectively allocate marketing resources for specific customer groups and to establish a healthy long-term relationship with customers.

## 2. Relevant Works

**2.1. RFM Model.** The RFM model was first proposed by Hughes [12]. As a popular tool of customer value analysis, it has been widely used for measuring customer lifetime value [13] and in customer segmentation and behavior analysis [14]. In the following paragraphs, we provide a brief description of the RFM model in the above literature.

RFM is short for recency, frequency, and monetary, which refer to recency of the last purchase, purchase frequency, and monetary value of purchase, respectively.  $R$  (recency) represents the time interval between a customer's last purchase date and the end date of a statistical period. The shorter the interval, the bigger the value of  $R$ .  $F$

(frequency) indicates the number of purchases made by the customer during the statistical period. The larger the value of  $F$  is, the higher the customer loyalty and the stronger the intention to repurchase would be.  $M$  (monetary) represents the total amount the customer spends in purchases during the statistical period. Hughes attached equal importance to these three variables [12], while Stone believed that the importance of the three variables varies among industries due to their different characteristics, suggesting unequal weights of these variables [13]. The RFM model is widely used in customer value analysis, and researchers have extended it according to different aspects.

Cheng and Chen combined RFM analysis with a rough set theory to establish rules for customer classification [13]. Chiang proposed an RFMDR model (based on an RFM/RFMD model), an extended version of RFM analysis, to identify valuable online shopping customers for the industry and to generate fuzzy association rules [15]. Kolarovszki et al. have proposed a novel modeling method for postal services using multidimensional segmentation. This CRM design proves useful in postal service companies [16]. Song et al. proposed a statistic-based approach to evaluate potential users via time series. With this approach, it is possible to segment time intervals of RFM in a large-scale dataset [17]. Aiming at the randomness of recent consumption time  $R$  index in the RFM model, and the colinearity between index  $F$  and index  $M$ , Bao et al. proposed an improved RFM model;  $K$ -means is used to cluster the customer data and proved the effectiveness of the improved RFM model [18]. In order to help the platform describe the multidimensional attributes of customers and more accurate customer value, an improved RFM model with five indicators is proposed. In this study, we use the RFM model as the basis to select the variables for clustering, so as to establish the clustering criteria objectively.

**2.2.  $K$ -Means++ Algorithm.** Clustering is the process of dividing a set of physical or abstract objects into groups of similar objects. The  $K$ -means algorithm, as one of the most popular clustering algorithms, was first used by MacQueen in 1967 [19], and it has been used extensively in various fields including data mining, statistical data analysis, and other business applications.

The literature survey reveals that one of the major applications of  $K$ -means is customer segmentation [20]. The  $K$ -means algorithm is widely used to effectively identify valuable customers and develop pertinent marketing strategies [21]. In particular, Cheng and Chen used the RFM model and  $K$ -means to perform customer relationship management, and experimental results demonstrate that the model they proposed is an effective method in customer value analysis [13].

$K$ -Means is a fast method in clustering analysis, but the accuracy and running time of clustering results largely depend on the location of the initial clustering center [22–24]. In order to solve the problem that  $K$ -means is sensitive to initial points, Arthur et al. proposed the  $K$ -means++ algorithm to improve the method of  $K$ -means randomly selecting initial clustering centers, that is, to make the

distance between the clustering centers as far as possible when selecting initial clustering centers. The results show that  $K$ -means++ can significantly improve the final error of classification results [25]. Because the  $K$ -means++ algorithm is accurate and efficient, this paper uses the  $K$ -means++ algorithm to classify customers.

### 3. Improved RFM Model and $K$ -Means++ Clustering Algorithm

In order to accurately identify the user value of the T-app community e-commerce platform, this study uses an improved RFM model to extract a user's features and uses the  $K$ -means++ clustering algorithm to achieve user classification. The indicators of the traditional RFM model portray customer characteristics from three perspectives: recency of the last purchase ( $R$ ), frequency of the purchases ( $F$ ), and monetary value of the purchases ( $M$ ). However, the user groups and scenarios studied in this article are quite different from the previous literature: (1) the user group is relatively fixed, (2) the consumer goods are relatively singular, and (3) the characteristics of repeated purchases are obvious. Therefore, based on the existing literature, we extract the characteristics of the users studied and improve and model the traditional indicators, as follows.

#### 3.1. Indicator Definition

- (1)  $R_1$ : average consumption time interval, refers to the average transaction time interval of a certain user in a certain period
- (2)  $F_1$ : frequency of the purchases, refers to the number of order transactions for a certain user within a certain period
- (3)  $M_1$ : monetary value of the purchases, refers to the total amount spent by a user in a certain period
- (4)  $S$ : customer contribution time, refers to the time interval from the first transaction to the last transaction of a certain user
- (5)  $P$ : repeat purchase attributes, refers to the number of purchases of a category of goods purchased by a certain user in the reference time

3.2. *Indicator Modeling.* The calculation formula of the  $R_1$  indicator is:

$$R_1 = \frac{T_{\text{last\_time}} - T_{\text{first\_time}}}{F_1}, \quad (1)$$

where  $T_{\text{last\_time}}$  indicates the time of the customer's last order transaction within the reference time period and  $T_{\text{first\_time}}$  indicates the time of the customer's first order transaction within the reference time period.

The calculation formula of the  $M_1$  indicator is

$$M_1 = \sum_i^n M_i, \quad (2)$$

where  $n$  represents the total number of a customer's consumption in the reference time period and  $M_i$  represents the amount of single consumption by the customer.

The calculation formula of the  $S$  indicator is

$$S = T_{\text{last\_time}} - T_{\text{first\_time}}. \quad (3)$$

In (3),  $T_{\text{first\_time}}$  is the time of the first transaction in the customer's purchase history.

In the improved RFM model, we adopted average consumption time interval ( $R_1$ ) to replace recency of the last purchase ( $R$ ) in the RFM model, thus overcoming the shortcomings of the large randomness of the  $R$  indicator in the traditional RFM model; for regular customers with high transaction frequency, the average order transaction time is more representative. The customer contribution time indicator ( $S$ ) reflects the customer's loyalty to the platform and continuous consumption ability, that is, the time interval from the first transaction in the customer's history to the last transaction at the reference time, while the repurchase indicator ( $P$ ) can describe the degree of customer dependence on a single product.

#### 3.3. Determination of Indicator Weights

3.3.1. *Step1-1: Standardization of Indicator Data.* Due to the different natures of the indicators in this model, the value of the average transaction interval  $R_1$  indicator is a negative indicator of customer value; the smaller the value of this indicator, the greater the customer value. On the contrary, the remaining four indicators are all positive indicators, that is, the greater the value of the customer under this indicator, the greater the customer value. Moreover, the five indicators of this model have inconsistent units and differ largely in value ranges. In order to achieve the purpose of normalization and standardization of heterogeneous indicators, the  $R_1$  indicator is specially standardized by the reverse standardization method, namely, formula (4), and the other indexes are standardized by the forward standardization method, namely, formula (5).

$$x'_{ij} = \frac{\max \{x_{ij}, \dots, x_{nj}\}}{\max \{x_{ij}, \dots, x_{nj}\} - \min \{x_{ij}, \dots, x_{nj}\}}, \quad (4)$$

$$x'_{ij} = \frac{x_{ij} - \min \{x_{ij}, \dots, x_{nj}\}}{\max \{x_{ij}, \dots, x_{nj}\} - \min \{x_{ij}, \dots, x_{nj}\}}. \quad (5)$$

In the above formula,  $x_{ij}$  represents the value of the  $j$ -th index of the  $i$ -th sample.

3.3.2. *Step1-2: Index Weight Calculation Using Entropy Method.* In order to obtain more objective segmentation results, in this study, we use the entropy weight method in

the objective weighting method to calculate the weights of the five indicators in the improved model.

### 3.4. Clustering Based on K-Means++ Algorithm

**3.4.1. Step2-1: Find the Optimal Number of K Clusters Using the Contour Coefficient Method.** Contour coefficient is an evaluation method of clustering effect, which was first proposed by Rousseeuw in 1986 [26]. It considers the agglomeration and separation of clusters comprehensively. Good clustering results should have both a smaller cohesion and a larger separation degree between clusters. In the research of this article, we use the maximum average contour coefficient method to determine the optimal cluster number  $K$ . The calculation process is as follows.

Suppose the data to be classified is divided into  $K$  clusters. For each vector in the cluster, calculate their contour coefficients separately. For one of the points  $i$ , the contour coefficient of the  $i$  vector is

$$S(i) = \frac{b(i) - a(i)}{\max \{a(i), b(i)\}}. \quad (6)$$

In formula (6),  $a(i)$  = average (the distance from the  $i$  vector to all other points in the cluster to which it belongs), expressed as the average value of the dissimilar degree from the  $i$  vector to other points in the same cluster;  $b(i)$  = min (the average distance from the  $i$  vector to all points in the nearest cluster), expressed as the minimum value of the average dissimilar degree from the  $i$  vector to other clusters. Average the contour coefficients of all vectors, and the result is the total contour coefficient of the clustering result. The larger the total contour coefficient value, the more ideal the clustering result.

The calculation result of the total profile coefficient is shown in Figure 1.

When  $K$  takes 4, the total contour coefficient of the cluster is 0.4022, which is the largest in the range of (3, 9). It can be concluded that the optimal cluster number  $K_{opt}$  is 4 in this data set.

**3.4.2. Step 2-2: Clustering.** The clustering algorithm process is shown in Figure 2.

In formula (a),  $D(x')^2 / \sum_{x \in X} D(x)^2$ , of Figure 2, value  $D(x)^2$  represents the distance from the data point  $x$  to the nearest cluster center (select  $S_i$  ( $S_i = x' \in X$ )). For formula (b) of Figure 2, the average error is equal to  $\sum_{i=1}^n [\min_{r=1, \dots, k} d(x_i, c_r)^2]$ .

## 4. Case Study

**4.1. Numerical Experiment.** The observation period is from September 1, 2018, to December 30, 2018, for three months. There are 3558 customer purchase records and 580 customers. The community shopping platform mainly sells 134 kinds of commodities such as cooked food and pasta. The data is processed as follows.

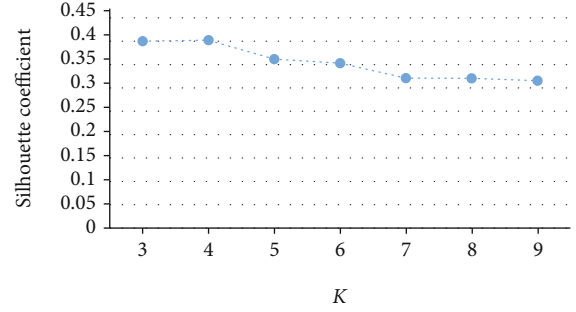


FIGURE 1: Comparison of contour coefficients when  $K$  is (3, 9).

**4.1.1. Raw Data Cleaning and Index Calculation.** The initial data consists of 12 dimensions such as user ID, product ID, purchase quantity, and consumption date. Four dimensions of user ID, consumption amount, consumption time, and product ID are selected from them, and the corresponding five indicators of each customer are calculated to form the initial data, as in Table 1.

In Table 2, the customer whose user ID is YH181102000001 has spent 52 days on the platform. A total of 5 purchases occurred during the reference time period, with a cumulative consumption of 64 yuan, and an average purchase was made on the platform every 10.40 days. The purchase of goods was relatively casual, and no repurchase of goods occurred (the value of  $P$  indicator is 1).

The  $R_1$  indicator adopts the reverse standardization method, namely, formula (4), and the other indicators adopt the forward standardization method, namely, formula (5) for standardization.

**4.1.2. Empower Indicators.** The entropy method is used to get the weight of each index. The weight of each index is shown in Table 2. At the same time, the weight of each index is multiplied by its corresponding weight to get the weighted data set.  $R_1'$ ,  $F_1'$ ,  $M_1'$ ,  $S'$ , and  $P'$  are the weighted indexes.

**4.2. Clustering Using K-Means++ Method.** After using the  $K$ -means algorithm to cluster the weighted data set, we get four different customer groups.

Exporting the standard data of four types of customer groups and calculating the average value of five indicators for each type of customer along with the number of users of each type of user, we get Table 3. The data of each user group in Table 3 is the average of the standardized data which has not been weighted.

**4.3. Customer Value Ranking and Value Analysis.** Since the user data after clustering at this time is weighted data, the total value of users of this type can be obtained by adding up the values of the indicators of each user category, calculated as follows:

$$\text{Value} = S' + F_1' + M_1' + R_1' + P'. \quad (7)$$

The calculated results are shown in Table 4.

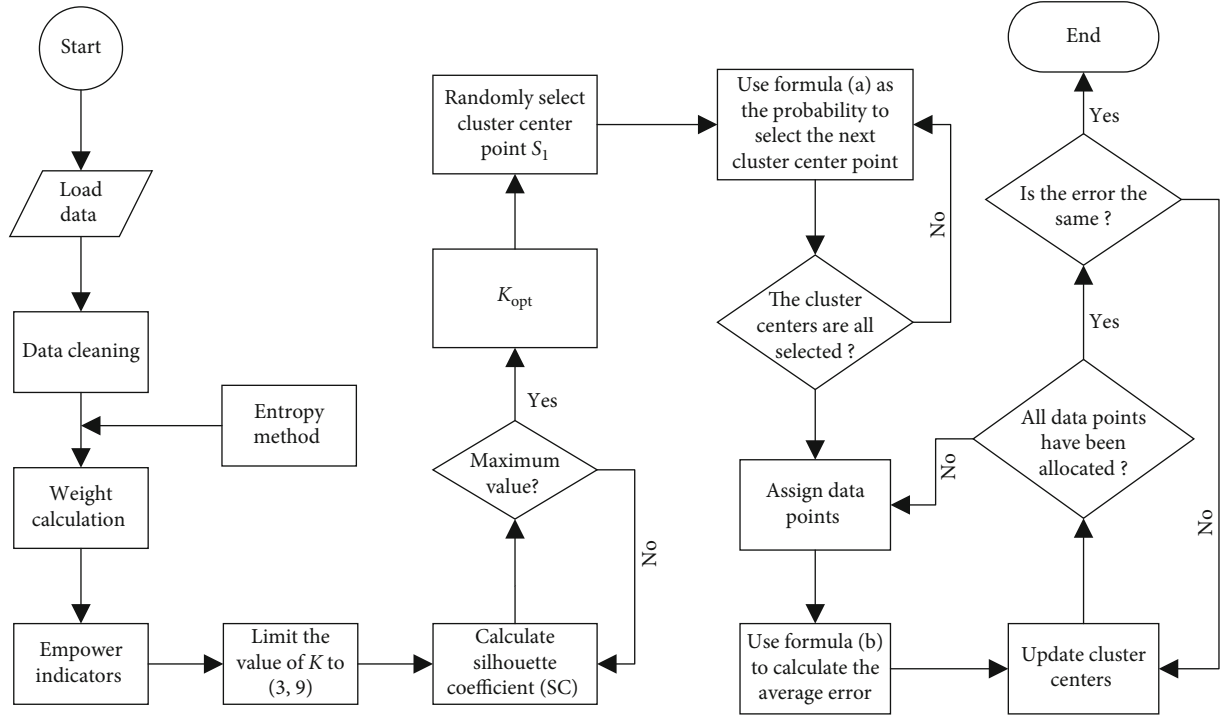


FIGURE 2: Algorithm flow chart.

TABLE 1: Part of customer data.

User ID	S (day)	$F_1$ (frequency)	$M_1$ (yuan)	$R_1$ (day)	P (frequency)
YH181102000001	52.0	5	64.0	10.40	1
YH170621000001	532.0	11	541.0	8.18	6

TABLE 2: Index weight.

Index	$R_1$	$F_1$	$M_1$	S	P
Weight	0.0044	0.2522	0.2752	0.1776	0.2903

TABLE 3: Customer clustering results.

Cluster	$R_1'$	$F_1'$	$M_1'$	$S'$	$P'$	Count
Cluster 3	0.00406	0.16857	0.09913	0.07957	0.1246	39
Cluster 2	0.00396	0.06675	0.04121	0.04815	0.0607	127
Cluster 4	0.00393	0.02145	0.01489	0.10469	0.0142	124
Cluster 1	0.00413	0.01295	0.00848	0.01855	0.0127	290

It can be seen from the model that the higher the amount of value, the greater the profit contribution of this type of customer to the platform. Therefore, according to the value ranking result and the different characteristics of the customer, we divided customers into important retention customers, development customers, loyal customers, and general customers. From the data, the characteristics of each type of customer are as follows.

4.3.1. *Type 1: Important Retention Customers.* Customer group 1 has the greatest value to the platform and is an important maintaining customer of the platform. Its S indicator is large among all customer groups, indicating that this customer group has maintained a long-term consumer relationship with the platform; the  $M_1$  indicator is much higher than other customer groups, indicating that this customer group has made a lot of consumption on the platform and is an important source of profit. At the same time, the value of the P indicator is also much higher than other customer groups, indicating that this customer group is more likely to repurchase a single product. The  $F_1$  indicator and the reverse-standardized  $R_1$  indicator indicate that this type of customer has greater consumer stickiness on the platform and purchases more frequently. As a result, the user characteristics of the platform's high-value customers can be characterized as follows: high consumption frequency, small consumption time interval, high degree of product specificity, and have their own fixed purchase product.

4.3.2. *Type 2: Developmental Customers.* Customer group 3 has a considerable performance in total consumption amount and consumption frequency. However, by observing its consumption time relationship with the platform,

TABLE 4: Customer value ranking.

Cluster	$R_1'$	$F_1'$	$M_1'$	$S'$	$P'$	Count	Quantity of value	Value ranking
Cluster 1	0.00406	0.16857	0.09913	0.07957	0.1246	39	0.475905	1
Cluster 3	0.00396	0.06675	0.04121	0.04815	0.0607	127	0.220726	2
Cluster 4	0.00393	0.02145	0.01489	0.10469	0.0142	124	0.159173	3
Cluster 2	0.00413	0.01295	0.00848	0.01855	0.0127	290	0.056788	4

it can be concluded that the users' time to contact the platform is average, and there is still room for tapping the value of its consumption. The platform should devote itself to transforming its development into important retention customers.

*4.3.3. Type 3: Loyal Customers.* Customer group 3 maintains the longest consumption relationship with the platform among all customers, indicating that this type of customer contacted the platform earlier and maintained a long-term consumption record. Although the consumption amount is not large, they give the platform's cash flow a greater guarantee, being loyal customers of the platform.

*4.3.4. Type 4: General Customers.* Customer group 2 does not have a long-term consumer relationship with the platform. The consumption frequency and the total amount of consumption are very low and so also is the repurchase of a single product, indicating that they had not made considerable consumption on the platform and had not contributed to the profit growth of the platform; thus, they belong to the general customers of the platform.

*4.4. Experimental Result Verification.* After analyzing the value of the customer, we return to the original data of the customer order transaction to verify the T-app customer value analysis result of the improved RFM model. The classification process of the traditional RFM model is similar with the model used in this article, that is, data standardization-calculation of weight-weighting-clustering-value ranking. Among them,  $F$  and  $M$  indicators use forward standardization and  $R$  indicators use reverse standardization. Later, according to the value ranking, the customers are also divided into important retention customers, loyal customers, development customers, and general customers. Now compare the classification results of the original RFM model with the improved RFM model classification results, as shown in Tables 5 and 6.

The analysis shows that under the RFM model, 34 important retention customers have been identified, all indicators of which are better than the rest of the customers, but no purchase behavior has been made recently. The consumer group 3 on the platform has a higher frequency of consumption and a larger amount of consumption, indicating that it has greater consumer stickiness and is a loyal customer of the platform. All indicators of customer group 4 consumption on the platform need to be improved, and they belong to developmental customers. Customer group 2 consumption amount and consumption frequency are both low, but the recent purchase

TABLE 5: The weight of each indicator of the RFM model.

Index	$R$	$F$	$M$
Weight	0.1601	0.4016	0.4382

time indicator has performed well, indicating that it may be due to the customers attracted by the platform's recent marketing strategy, which means that they did not bring considerable profits and cash flow to the platform. So this group belongs to general customers.

The comparison shows that the original RFM model cannot identify the new and old attributes of customers and can only determine the customer's loyalty index based on the customer's consumption frequency and consumption amount, but this has a certain contingency. Customers who make high-frequency consumption in the short-term may terminate their purchases after the platform's promotional activities are over. Such customers should not be considered as having high long-term value. At the same time, improving the RFM model can intuitively feel the differences in various indicators of different user groups. For example, users who spend high on the platform have high repurchase of goods, which can guide the platform to carry out differentiated marketing and improve the customers' experience and ultimately bring platform profit growth.

Now take the customer value indicators and customer classification results of some customers through the RFM model and the improved RFM model as an example to illustrate the credibility of the improved RFM model, as shown in Tables 7 and 8.

In Tables 7 and 8, we can see that the user with the ID of YH171201000030 has the same classification results under the two models. Analysis of the RFM model shows that the user of YH171201000030 made a total of 529 yuan worth of purchases during the observation period, with a frequency of 24 purchases, and the recent purchase time  $R$  index is relatively close, indicating that the customer has a greater probability of repurchasing next. But at the same time, it can also be found that the RFM model cannot accurately identify whether the user has a greater degree of dependence on a single product and the average time interval for placing orders nor can it define loyalty attributes based on the relationship time between the customer and the platform. For this type of important retention customers, improving the RFM model can provide a more detailed description based on T-app user data, which facilitates the platform to conduct one-to-one precision marketing, which is conducive to the retention of such customers.

TABLE 6: RFM model customer segmentation value analysis.

Cluster	$R'$	$F'$	$M'$	Count	Value ranking
Cluster 1	0.1474	0.1622	0.2849	34	1
Cluster 3	0.1135	0.0792	0.1202	106	2
Cluster 4	0.1165	0.0192	0.0324	188	3
Cluster 2	0.0293	0.0152	0.0225	252	4

TABLE 7: RFM model user classification results (partial).

User ID	$R$	$F$	$M$	Customer segmentation results
YH171201000030	2.0	24	529.0	Important retention customers
YH181207000001	19.0	2	14.0	Development client
YH181018000002	72.0	8	138.5	Loyal customers

TABLE 8: User classification results of the improved RFM model (partial).

User ID	$R_1$	$F_1$	$M_1$	$S$	$P$	Customer segmentation results
YH171201000030	4.75	24	529.0	380.0	8	Important retention customers
YH181207000001	1.5	2	14.0	3.0	1	General customer
YH181018000002	0	8	138.5	0	5	General customer

For the user with the ID “YH181207000001” under the RFM model, we only know that he had made a total amount of 14 yuan and a frequency of 2, but because his latest purchase happened to be closer to the end of the latest purchase period, he was classified as a development customer. While under the improved RFM model, it can be known that the user has only been exposed to the platform for 3 days and has only made two low-value consumptions on the platform, and his potential commercial value needs to be further developed, so he is correctly divided into general customers. Therefore, the improved RFM model can overcome the randomness of the  $R$  indicator in the original RFM model and can accurately locate such new customers.

The user with the ID “YH181018000002” made 8 purchases with a total amount of 138.5 yuan on the platform, and the last purchase was 72 days away from the end of the reference period. Because of his relatively considerable consumption data in the overall sample, he was classified as a loyal customer in the original RFM model. However, through the analysis of the improved RFM model, it can be known that the user has considerable data on the total amount of consumption and consumption frequency. But he was concentrated on a certain product 5 times in a day and had not contacted the platform, indicating that he may be attracted by the platform’s recent promotion activities, so the improved RFM model has correctly divide him into general customers, waiting for further observation.

The experimental results given above can show that the improved RFM model can provide a more accurate user description than the original RFM model under the same classification. Moreover, it overcomes the randomness of the  $R$  indicator in the original RFM model and perfected

the RFM model’s shortcomings of only describing the customer stickiness and loyalty through observation of consumption amount and frequency. The improved model can better complete the user value evaluation of T-app.

## 5. Conclusions

In order to accurately analyze the customer value of a T-app community e-commerce platform, a customer value analysis method based on an improved RFM model is proposed. The average order transaction time interval, customer transaction times in a certain period, customer total consumption amount in a certain period, customer relationship duration, five repurchase indicators, use of forward and reverse standardization methods to standardize the indicators, and use of the objective assignment method of information theory entropy method to calculate the weight of the five indicators are included. The concept of the silhouette coefficient is used to determine the best  $K$  value, and the  $K$ -means++ clustering algorithm is used to cluster the weighted indicators; finally, customers are divided into different value customer groups. This paper takes the T-app order transaction data from September 1, 2018, to December 30, 2018, to analyze the customer value, and the results show that the improved RFM model can more accurately analyze the customer value of T-app.

## Data Availability

All information is within the article.



## Conflicts of Interest

The authors have declared that no competing interests exist concerning this study.

## Acknowledgments

This work was supported, in part, by the Funds for the Beijing University of Chemical Technology First-Class Discipline Construction (XK1802-5) and Beijing University of Chemical Technology (GJD202002).

## References

- [1] A. M. Al-Momani, M. A. Mahmoud, and M. S. Ahmad, "Factors that influence the acceptance of internet of things services by customers of telecommunication companies in Jordan," *Journal of Organizational and End User Computing*, vol. 30, no. 4, pp. 51–63, 2018.
- [2] I. Kitouni, D. Benmerzoug, and F. Lezzar, "Smart agricultural enterprise system based on integration of internet of things and agent technology," *Journal of Organizational and End User Computing*, vol. 30, no. 4, pp. 64–82, 2018.
- [3] K. G. Srinivasa, B. J. Sowmya, A. Shikhar, R. Utkarsha, and A. Singh, "Data analytics assisted internet of things towards building intelligent healthcare monitoring systems," *Journal of Organizational and End User Computing*, vol. 30, no. 4, pp. 83–103, 2018.
- [4] S. K. Biswas, D. Devi, and M. Chakraborty, "A hybrid case based reasoning model for classification in internet of things (IoT) environment," *Journal of Organizational and End User Computing*, vol. 30, no. 4, pp. 104–122, 2018.
- [5] M. R. Reddy, K. G. Srinivasa, and B. E. Reddy, "Smart vehicular system based on the internet of things," *Journal of Organizational and End User Computing*, vol. 30, no. 3, pp. 45–62, 2018.
- [6] R. Parada, J. Melià-Seguí, and R. Pous, "Anomaly detection using RFID-based information management in an IoT context," *Journal of Organizational and End User Computing*, vol. 30, no. 3, pp. 1–23, 2018.
- [7] R. Heldt, C. S. Silveira, and F. B. Luce, "Predicting customer value per product: from RFM to RFM/P," *Journal of Business Research*, vol. 3, 2019.
- [8] A. Griva, C. Bardaki, K. Pramatar, and D. Papakiriakopoulos, "Retail business analytics: customer visit segmentation using market basket data," *Expert Systems with Applications*, vol. 100, pp. 1–16, 2018.
- [9] K. Khalili-Damghani, F. Abdi, and S. Abolmakarem, "Hybrid soft computing approach based on clustering, rule mining, and decision tree analysis for customer segmentation problem: real case of customer-centric industries," *Applied Soft Computing*, vol. 73, pp. 816–828, 2018.
- [10] H.-C. Chang, "Developing EL-RFM model for quantification learner's learning behavior in distance learning," in *2010 2nd International Conference on Education Technology and Computer*, Shanghai, China, 2010.
- [11] Y. S. Cho, S. C. Moon, S. C. Noh, and K. H. Ryu, "Implementation of personalized recommendation system using  $k$ -means clustering of item category based on RFM," in *In 2012 IEEE International Conference on Management of Innovation & Technology*, pp. 378–383, Bali, Indonesia, 2012.
- [12] A. M. Hughes, *Strategic Database Marketing*, Probus Publishing Company, Chicago, USA, 1994.
- [13] C. H. Cheng and Y.-S. Chen, "Classifying the segmentation of customer value via RFM model and RS theory," *Expert Systems with Applications*, vol. 36, no. 3, pp. 4176–4184, 2009.
- [14] D. Chen, S. L. Sain, and K. Guo, "Data mining for the online retail industry: a case study of RFM model-based customer segmentation using data mining," *Journal of Database Marketing & Customer Strategy Management*, vol. 19, no. 3, pp. 197–208, 2012.
- [15] W. Y. Chiang, "To mine association rules of customer values via a data mining procedure with improved model: an empirical case study," *Expert Systems with Applications*, vol. 38, no. 3, pp. 1716–1722, 2011.
- [16] P. Kolarovszki, J. Tengler, and M. Majerčáková, "The new model of customer segmentation in postal enterprises," *Procedia - Social and Behavioral Sciences*, vol. 230, pp. 121–127, 2016.
- [17] M. Song, X. Zhao, E. Haihong, and Z. Ou, "Statistics-based CRM approach via time series segmenting RFM on large scale data," *Knowledge-Based Systems*, vol. 132, no. 15, pp. 282–291, 2017.
- [18] Z. Bao, Y. Zhao, and Y. Li, "Analysis on Baidu take-way customer value based on improved RFM model," *Journal of Xi'an University of Posts and Telecommunications*, vol. 24, no. 1, pp. 105–110, 2019.
- [19] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, no. 14, pp. 281–297, 1967.
- [20] D. Arunachalam and N. Kumar, "Benefit-based consumer segmentation and performance evaluation of clustering approaches: an evidence of data-driven decision-making," *Expert Systems with Applications*, vol. 111, pp. 11–34, 2018.
- [21] A. Mesforoush and M. J. Tarokh, "Customer profitability segmentation for SMEs case study: network equipment company," *International Journal of Research in Industrial Engineering*, vol. 2, no. 1, pp. 30–44, 2013.
- [22] A. K. Jain, "Data clustering: 50 years beyond  $K$ -means," *Pattern Recognition Letters*, vol. 31, no. 8, pp. 651–666, 2010.
- [23] D. Steinley, "K-means clustering: a half-century synthesis," *British Journal of Mathematical and Statistical Psychology*, vol. 59, no. 1, pp. 1–34, 2006.
- [24] J. Wu, L. Shi, W. P. Lin et al., "An empirical study on customer segmentation by purchase behaviors using a RFM model and  $K$ -means algorithm," *Mathematical Problems in Engineering*, vol. 2020, Article ID 8884227, 7 pages, 2020.
- [25] D. Arthur and S. Vassilvitskii, *k-Means++: the advantages of careful seeding*, Stanford, 2006.
- [26] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of computational and applied mathematics*, vol. 20, pp. 53–65, 1987.

## Research Article

# Entropy-Based Multiview Data Clustering Analysis in the Era of Industry 4.0

Yi Gu  and Kang Li

*School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi, Jiangsu 214122, China*

Correspondence should be addressed to Yi Gu; 8202101437@jiangnan.edu.cn

Received 30 March 2021; Revised 16 April 2021; Accepted 19 April 2021; Published 3 May 2021

Academic Editor: Shan Zhong

Copyright © 2021 Yi Gu and Kang Li. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In the era of Industry 4.0, single-view clustering algorithm is difficult to play a role in the face of complex data, i.e., multiview data. In recent years, an extension of the traditional single-view clustering is multiview clustering technology, which is becoming more and more popular. Although the multiview clustering algorithm has better effectiveness than the single-view clustering algorithm, almost all the current multiview clustering algorithms usually have two weaknesses as follows. (1) The current multiview collaborative clustering strategy lacks theoretical support. (2) The weight of each view is averaged. To solve the above-mentioned problems, we used the Havrda-Charvat entropy and fuzzy index to construct a new collaborative multiview fuzzy c-means clustering algorithm using fuzzy weighting called Co-MVFCM. The corresponding results show that the Co-MVFCM has the best clustering performance among all the comparison clustering algorithms.

## 1. Introduction

In the era of Industry 4.0, as the methods of data collection become more and more diverse, the complexity of data is also increasing. For example, a driverless car will collect environmental data through a variety of sensors and conduct analysis and processing from multiple views while driving. In unsupervised learning, clustering is usually used for complexity data analysis. However, traditional clustering methods, such as K-means [1, 2], fuzzy C-means (FCM) [3, 4], maximum entropy clustering (MEC) [5, 6], and possibilistic C-means (PCM) [7, 8], are all designed for single-view data analysis. When the single-view algorithms [9–11] encounter a multiview clustering task, the common practice is to first consider each view independently, treating each view as an independent clustering task. After finishing each clustering task of each view, the integrated learning mechanism [12] is used to select an appropriate integrated learning strategy to integrate multiple clustering results and then get the final cluster-

ing results. However, due to obvious deviation of clustering results in a certain view or great difference of clustering results among different views, the multiview strategy which artificially separates each view for independent analysis may result in inaccurate global partitioning results obtained by integrated learning or unstable algorithm performance.

In many real applications, multiview representation of data is becoming more and more popular [13], especially in the field of medicine [14, 15]. For example, people's living standard and economic situation have been improved since entering the twentieth century. But the incidence rate of cancer has increased by nearly 50% compared with that of the 1880s due to environmental pollution, food safety, and working pressure. However, with the continuous improvement of medical level, various detection methods such as laboratory examination (routine examination, serological examination, gene or gene product examination, etc.), imaging and endoscopy (X-ray examination, B-ultrasound examination, CT examination, radionuclide imaging, etc.), and

cytopathological examination (puncture biopsy, forceps biopsy, section analysis, etc.) have been proposed and applied. These methods can be used to analyze the suspected patients from different views. This is a typical multiview data representation problem. The above examples reveal that developing various multiview clustering algorithms is very necessary for us to better observe and mine the essence of data from the viewpoint of its diverse descriptions and accordingly obtain a better clustering result that simultaneously satisfies every representation (view). Introducing the multiview technology into the traditional clustering analysis method so that there is collaborative learning in the clustering process is considered to be an effective solution. In recent years, some effective multiview clustering methods have been proposed using the above strategies. Yamanishi et al. [16] proposed a collaborative clustering algorithm Co-EM algorithm that can be used to solve multiview problems based on the EM algorithm from the perspective of probability and test the proposed algorithm's effectiveness through some text-like samples. Inspired by FCM algorithm, Pedrycz [17] controlled the fuzzy partition between the various views, constructed a divisional cooperative control function, and finally obtained the Co-FC algorithm. The algorithm has shown certain advantages on various datasets. More early related multiview clustering studies can be found in [18, 19].

As mentioned above, a lot of researches have begun to focus on the construction of multiview clustering algorithm. Through the summary of the current research on multiview clustering, we find that the current research mainly focuses on the following aspects: (1) the early multiview clustering algorithm usually preprocesses the data itself, and the most direct method is to synthesize a multiview data into a single-view data through feature fusion and then use the data clustering analysis; (2) most of the multiview clustering algorithms proposed in recent years use collaborative learning strategy, which can enhance the performance of each view data in the process of clustering; (3) when most multiview clustering algorithms with collaborative learning ability treat each view, their common practice is to average the weight of each view. In particular, we take the most classic Co-FKM algorithm [20] as an example, which is one of the representative multiview clustering algorithms in recent years. The algorithm designs a very effective multiperspective collaborative learning strategy, which can make the data between different perspectives use membership to complete collaborative learning in the process of clustering. But the algorithm also has a fatal disadvantage, that is, it just treats each perspective equally and does not give each perspective different weights. In addition, the multiperspective collaborative learning strategy proposed by the algorithm also lacks the necessary physical meaning, which cannot explain why this collaborative learning strategy can contribute to the final clustering performance. In response to the above-mentioned challenges, in this study, a new view space division criterion is first constructed based on the Havrda-Charvat entropy, which is used to control the space division results across different views so that the space division results of each view tend to be as consistent as possible in order to obtain a more stable and more

comprehensive global spatial division result. Furthermore, we introduced fuzzy index and fuzzy weights to adaptively weight each view and effectively adjust the weight of each view during the clustering process so that the view with the clearest spatial division has a larger weight. Finally, a new collaborative multiview fuzzy c-means clustering algorithm using fuzzy weighting called Co-MVFCM is proposed by combining with the Havrda-Charvat entropy and fuzzy index. We summarized the contributions of this study here:

- (1) We construct a new view space division criterion using the Havrda-Charvat entropy. The built criterion can be used to control the space division results across different views
- (2) We construct a view weighting mechanism using fuzzy index. The new view weighting mechanism can be used to recognize the importance degree of each view

Overall, our proposed Co-MVFCM algorithm not only has good space division ability but also has the ability to adaptively recognize the best view.

## 2. Related Work

When multiview clustering task is coming, Cleuziou et al. [20] proposed the Co-FKM method based on classical FCM. In Co-FKM, multiview clustering is achieved by a constraint of fuzzy membership degree which is aimed at keeping the partition result of each view as consistent as possible. Here, the Co-FKM method is defined as

$$J = \sum_{r=1}^R \sum_{i=1}^C \sum_{j=1}^N \left[ \mu_{ij,r}^m \|x_{j,r} - v_{i,r}\|^2 + \lambda \Delta_r \right], \quad (1)$$

$$\Delta_r = \lambda \frac{1}{R-1} \sum_{r'=1, r' \neq r}^R \sum_{i=1}^C \sum_{j=1}^N \left( \mu_{ij,r'}^m - \mu_{ij,r}^m \right) \|x_{j,r} - v_{i,r}\|^2 \quad (2)$$

$$\text{s.t.} \quad \sum_{i=1}^C \mu_{ij,r} = 1,$$

and  $\mu_{ij,r} \in [0, 1]$ ,  $1 \leq j \leq N$ ,  $1 \leq r \leq R$ .

By substituting Equation (2) to Equation (1), Equation (1) is simplified as

$$J = \sum_{r=1}^R \sum_{i=1}^C \sum_{j=1}^N \left[ \tilde{\mu}_{ij,r,\lambda} \|x_{j,r} - v_{i,r}\|^2 \right], \quad (3)$$

where  $\tilde{\mu}_{ij,r,\lambda} = (1 - \lambda)\mu_{ij,r}^m + (\lambda/(R-1))\sum_{r'=1, r' \neq r}^R \mu_{ij,r'}^m$ ,  $\lambda$  can be used to control the contribution of  $\Delta_r$ , and  $\tilde{\mu}_{ij,r,\lambda}$  is the average fuzzy membership degree of  $\mu_{ij,r}^m$  and  $\mu_{ij,r'}^m$ .

The objective function of Co-FKM can be optimized by introducing Lagrange multipliers. So the fuzzy membership degree  $\mu_{ij,r}$  and center  $v_{i,r}$  are obtained as

$$v_{i,r} = \frac{\sum_{j=1}^N \tilde{\mu}_{ij,r,\lambda} x_{j,r}}{\sum_{j=1}^N \tilde{\mu}_{ij,r,\lambda}}, \quad (4)$$

$$\mu_{ij,r} = \frac{1}{\sum_{h=1}^C \left[ \left( (1-\lambda)d_{ij,r}^2 + (\lambda/(R-1)) \sum_{r'=1, r' \neq r}^R d_{ij,r'}^2 \right) / \left( (1-\lambda)d_{hj,r}^2 + (\lambda/(R-1)) \sum_{r'=1, r' \neq r}^R d_{hj,r'}^2 \right) \right]^{1/(m-1)}}.$$

To obtain a fuzzy division standard with global considerations, the fuzzy membership of each view can be computed by using the geometric mean method [20]. The specific expression is as follows:

$$\hat{\mu}_{ij} = R \sqrt{\prod_{r \in R} \mu_{ij,r}}. \quad (5)$$

From the Co-FKM algorithm, we can draw a general framework to represent multiview clustering, which is illustrated in Figure 1. The Co-FKM algorithm incorporates the spatial division and approximation criteria across different views in the clustering and realizes collaborative learning across different views; it has more effective multiview clustering performance compared with traditional single-view integrated clustering technology. However, as we stated before, it still has challenges to be further solved.

### 3. Collaborative Multiview Fuzzy Clustering (Co-MVFCM) Using Entropy Technology

In view of the two shortcomings of the current multiview clustering methods, the following two new technologies based on the entropy theory were introduced.

- (1) We use the Havrda-Charvat entropy to construct a new view space division approximation criterion and find the maximum similarity component between each view so that while improving the performance of clustering, it also gives the view space division approximation criterion new physical meaning from the perspective of entropy
- (2) We propose an entropy-weighted multiview clustering technology. By weighting each view, we can find the best view in the iterative optimization process and get the best fuzzy division result at the same time, in order to effectively control the weight

Figure 2 illustrates the new framework of multiview clustering.

**3.1. Approximation Criterion of Space Division from Different Views Based on the Havrda-Charvat Entropy.** In this study, the Havrda-Charvat entropy of  $x_{j,r}$  is defined as

$$\Psi^{(m)}(x_{j,r}) = \frac{1}{1 - 2^{1-m}} \sum_{i=1}^C (\mu_{ij,r}^m - 1). \quad (6)$$

It is obvious that if the fuzzy membership degree  $\mathbf{U}_r = [\mu_{ij,r}^m]$  is considered as a probability matrix, when the constraint  $\sum_{i=1}^C \mu_{ij,r} = 1$  holds,  $\Psi^{(m)}(x_{j,r})$  is equal to 0. It is very intuitive to show that the uncertainty of  $x_{j,r}$  belonging to each division in the sample set  $\mathbf{X}_r$  of this view reaches the minimum value. That is to say, when the objective function reaches its minimum value, the Havrda-Charvat entropy of  $\mu_{ij,r}$  also reaches its minimum value.

Although Equation (6) can ensure that the uncertainty of division can be minimized, it is limited to a single view. In order to expand it into a field of multiple views, in this study, we expand Equation (6) into the following expression form by referring to the relevant strategies used in [20]:

$$\Phi = \frac{1}{1 - 2^{1-m}} \sum_{i=1}^C \sum_{j=1}^N (\mu_{ij,r}^m - 1) + \beta \frac{1}{1 - 2^{1-m}} \frac{1}{R-1} \sum_{i=1}^C \sum_{j=1}^N \sum_{r'=1, r' \neq r}^R (\mu_{ij,r'}^m - 1) \cdot \left[ (\mu_{ij,r'}^m - 1) - (\mu_{ij,r}^m - 1) \right]. \quad (7)$$

So we have

$$\Phi = (1 - \beta) \frac{1}{1 - 2^{1-m}} \sum_{i=1}^C \sum_{j=1}^N (\mu_{ij,r}^m - 1) + \beta \frac{1}{1 - 2^{1-m}} \frac{1}{K-1} \sum_{r'=1, r' \neq r}^R \sum_{i=1}^C \sum_{j=1}^N (\mu_{ij,r'}^m - 1). \quad (8)$$

We observe from Equations (7) and (8) that  $\eta$  can be used to effectively regulate the weight relationship between the current view and the membership degree division of other views ( $0 < \beta < 1$  and  $\beta = (R-1)/R$ ). So we can get the weighted average of membership degree and finally make the membership degree division of each view as consistent as possible, so as to obtain the spatial division result with a more global view.

**3.2. Multiview Adaptive Weighting Based on Fuzzy Index.** In this study, we develop an automatic view weighting strategy using fuzzy index to recognize the best view. Suppose  $w_r$ ,

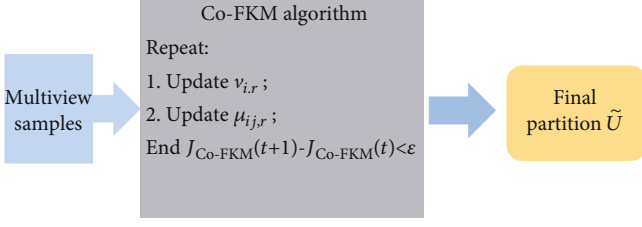


FIGURE 1: The principle of Co-FKM algorithm.

represents the weight of view  $r$  under the condition that  $\sum_{r=1}^R w_r = 1$  and  $w_r \geq 0$ , then  $w_r$  can be considered as the probability distribution which is defined as

$$f(w_r) = \sum_{r=1}^R w_r^p. \quad (9)$$

Fuzzy index technology is introduced through the above methods to make the objective function achieve the optimal entropy as much as possible, which is also the classical fuzzy c-means clustering principle [4].

3.3. *Co-MVFCM*. According to the above definitions, we propose our new multiview clustering method. The objective function of Co-MVFCM is

$$J_{\text{Co-MVFCM}} = \sum_{r=1}^R w_r^p \left[ \sum_{i=1}^C \sum_{j=1}^N \mu_{ij,r}^m \|x_{j,r} - v_{i,r}\|^2 + \lambda \Phi_r \right],$$

$$\Phi_r = (1 - \beta) \frac{1}{1 - 2^{1-m}} \sum_{i=1}^C \sum_{j=1}^N (\mu_{ij,r}^m - 1)$$

$$+ \beta \frac{1}{1 - 2^{1-m}} \frac{1}{R-1} \sum_{r'=1, r' \neq r}^R \sum_{i=1}^C \sum_{j=1}^N (\mu_{ij,r'}^m - 1) \quad (10)$$

$$\text{s.t. } \mu_{ij,r} \in [0, 1],$$

$$\sum_{i=1}^C \mu_{ij,r} = 1 \text{ and } \sum_{r=1}^R w_r = 1, 1 \leq r \leq R, 1 \leq j \leq N.$$

Obviously, the objective function contains two main parts. The first one is  $\ell(\mathbf{U}, \mathbf{V}, \mathbf{X}) = \sum_{i=1}^C \sum_{j=1}^N \mu_{ij,r}^m \|x_{j,r} - v_{i,r}\|^2 + \lambda \Phi_r$  which is derived from Havrda-Charvat and used for collaborative clustering. The essence of the first part is to find out as many similar parts among different perspectives as

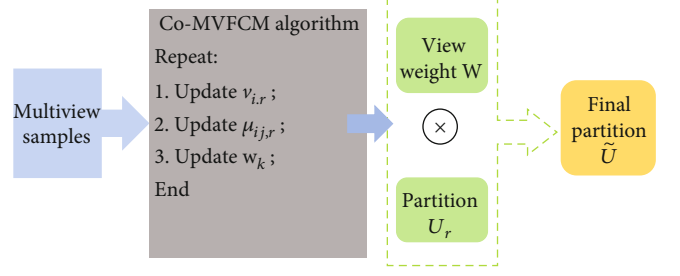


FIGURE 2: The principle of our proposed Co-MVFCM method.

possible through multiview clustering technology and finally make the spatial division results of different views tend to be the same. The second part is  $\tilde{\ell}(\mathbf{U}, \mathbf{V}, \mathbf{X}, \mathbf{W}) = \sum_{r=1}^R w_r^p \ell(\mathbf{U}, \mathbf{V}, \mathbf{X})$  which is derived from fuzzy index. This part can be used to adaptively calculate the weight values of each view, and finally, when the algorithm reaches the optimal level, the optimal view partitioning results can be obtained according to the weight matrix of the views. The parameter  $\beta$  can be set to  $\beta = (R - 1)/R$ . The parameter  $\lambda$  can be determined by using grid optimization [13, 21].

To obtain the final result of space division with global characteristics, the integration strategy of global space division mentioned in [20] is abandoned in this study. We define a new integration strategy to obtain the final space division as

$$\tilde{\mathbf{U}} = \sum_{r=1}^R w_r^R \sqrt{\mathbf{U}_r}. \quad (11)$$

3.3.1. *Optimization*. The proposed multiview can be optimized by introducing Lagrange multipliers. In this section, we give three theorems to obtain updating rules in terms of fuzzy membership degree, view weights, and cluster centers.

**Theorem 1.** When  $\mathbf{U}_r$  and  $\mathbf{W}_r$  are fixed, the cluster center can be solved by

$$\mathbf{v}_{i,r} = \frac{\sum_{j=1}^N \mu_{ij,r}^m \mathbf{x}_{j,r}}{\sum_{j=1}^N \mu_{ij,r}^m}. \quad (12)$$

*Proof.* By setting  $\partial J / \partial V_r = 0$ , we have  $\mathbf{v}_{i,r} = \sum_{j=1}^N \mu_{ij,r}^m \mathbf{x}_{j,r} / \sum_{j=1}^N \mu_{ij,r}^m$ . Therefore, Theorem 1 is achieved.

**Theorem 2.** When the cluster center  $\mathbf{V}_r$  and view weight matrix  $\mathbf{W}_r$  are fixed, the fuzzy membership degree matrix can be solved by

$$\mu_{ij,r} = \frac{1}{\left[ \sum_{h=1}^C ((1 - \beta) / (1 - 2^{1-m})) \lambda + \|\mathbf{x}_{j,r} - \mathbf{v}_{i,r}\|^2 / ((1 - \beta) / (1 - 2^{1-m})) \lambda + \|\mathbf{x}_{j,r} - \mathbf{v}_{h,r}\|^2 \right]^{1/(m-1)}}. \quad (13)$$

*Proof.* By introducing Lagrange multipliers  $\alpha_{j,r}$  and considering the constraint  $\sum_{i=1}^C \mu_{ij,r} = 1$ , we have the following objective function:

$$\begin{aligned} J_{\text{Co-MVFCM}}(\mu_{ij,r}, \beta_{j,r}) &= - \sum_{r=1}^R \sum_{j=1}^N o_{j,r} \left( \sum_{i=1}^C \mu_{ij,r} - 1 \right) \\ &\quad + \sum_{r=1}^R w_r^p \left[ \sum_{i=1}^C \sum_{j=1}^N \mu_{ij,r}^m \|x_{j,r} - v_{i,r}\|^2 + \lambda \Phi_r \right], \end{aligned} \quad (14)$$

where  $\Phi_r = \beta(1/(1-2^{1-m}))(1/(R-1)) \sum_{r'=1, r' \neq r}^R \sum_{i=1}^C \sum_{j=1}^N (\mu_{ij,r'}^m - 1) + (1-\beta)(1/(1-2^{1-m})) \sum_{i=1}^C \sum_{j=1}^N (\mu_{ij,r}^m - 1)$ .

$$\mu_{ij,r} = \frac{1}{\left[ \sum_{h=1}^C \|x_{j,r} - v_{h,r}\|^2 + ((1-\beta)/(1-2^{1-m}))\lambda \|x_{j,r} - v_{h,r}\|^2 + ((1-\beta)/(1-2^{1-m}))\lambda \right]^{1/(m-1)}}. \quad (17)$$

Therefore, the proof of Theorem 2 is achieved.

**Theorem 3.** When the center matrix  $\mathbf{V}_r$  and the fuzzy membership degree matrix  $\mathbf{U}_r$  are fixed, the weight matrix can be solved by

$$w_r = \frac{1}{\left[ \sum_{r'=1}^R \sum_{i=1}^C \sum_{j=1}^N \mu_{ij,r'}^m \|x_{j,r} - v_{i,r}\|^2 + \lambda \Phi_r / \sum_{i=1}^C \sum_{j=1}^N \mu_{ij,r'}^m \|x_{j,r'} - v_{i,r'}\|^2 + \lambda \Phi_{r'} \right]^{1/(p-1)}}. \quad (18)$$

*Proof.* Similar to Theorem 2, by introducing Lagrange multipliers and considering the constraint  $\sum_{r=1}^R w_r = 1$ , we have Theorem 3.

*Remark 4.* In this section, a novel multiview clustering method called Co-MVFCM is proposed. The proposed Co-MVFCM method can find the most important view adaptively, and it also can obtain the best space division by using Equation (11). However, we will find that the proposed Co-MVFCM method has three predefined parameters. These predefined parameters should be defined by using grid optimization which will lose many time costs. In the near future, we will consider how to reduce the number of these predefined parameters.

## 4. Experimental Studies

*4.1. Settings.* In this study, we introduce several UCI datasets to evaluate the proposed multiview clustering method. For

By setting the partial derivative of  $J_{\text{Co-MVFCM}}(\mu_{ij,r}, o_{j,r})$  w.r.t.  $\mu_{ij,r}, o_{j,r}$  to 0, i.e.,  $\partial J_{\text{Co-MVFCM}} / \partial \mu_{ij,r} = 0$ , we have

$$\mu_{ij,r} = \left[ \frac{o_{j,r}}{mw_r^p \left( ((1-\beta)/(1-2^{1-m}))\lambda + \|x_{j,r} - v_{i,r}\|^2 \right)} \right]^{1/(m-1)}. \quad (15)$$

Similarly, with  $\partial J_{\text{Co-MVFCM}} / \partial o_{j,r} = 0$ , we have

$$\sum_{i=1}^C \mu_{ij,r} = 1. \quad (16)$$

By combining Equations (15) and (16) to remove  $o_{j,r}$ , we have

fair comparison, Co-FKM [20], LSSMTC [22], CombKM [22], and Coclustering [23] are introduced for benchmarking testing.

We introduce two commonly used criteria, i.e., NMI and RI to evaluate all clustering methods. They are defined as follows.

- (1) Normalized Mutual Information (NMI) [24, 25]

$$\text{NMI} = \frac{\sum_{i=1}^C \sum_{j=1}^C N_{i,j} \log N \cdot N_{i,j} / N_i \cdot N_j}{\sqrt{\sum_{i=1}^C N_i \log N_i / N \cdot \sum_{j=1}^C N_j \log N_j / N}}, \quad (19)$$

where  $N_i$  represents the number of samples in the  $i$ th cluster,  $N_{i,j}$  represent the matching degree of the  $i$ th cluster and the  $j$ th cluster, and  $N$  represents the size of the dataset.

INPUT: Multiview dataset  $\text{Data}_r = \{\mathbf{X}_1, \dots, \mathbf{X}_N\}$ . Number of clusters  $C(2 \leq C < N)$ , stop threshold  $\epsilon$ , fuzzy index  $m$  and  $p$ , number of iteration  $f$ , maximum number of iteration  $L$ , regularization parameter  $\lambda$   
 OUTPUT: Center of each cluster  $\mathbf{v}_{i,r}$ , global fuzzy membership degree matrix  $\mathbf{U}$ , view weight  $w_r$   
 Step 1. Randomize  $\mathbf{v}_i$ ,  $\mu_{ij}$ , and  $w_r$   
 Step 2. Update  $\mathbf{v}_{i,r}$  based on Equation (12)  
 Step 3. Update  $\mu_{ij,r}$  based on Equation (13)  
 Step 4. Update  $w_r$  based on Equation (18)  
 Step 5. If  $\|J^{l+1} - J^l\| < \epsilon$  or  $f > L$ , stop loop; otherwise, go to Step 2  
 Step 6. Compute  $\tilde{\mathbf{U}}$  based on Equation (11)

ALGORITHM 1: Descriptions of the Co-MVFCM method.

TABLE 1: The description of some UCI datasets.

Datasets	Description of these datasets	The number of the dimensions	The number of the clusters	The number of the samples
Iris	Classes of <i>Iris</i> plants	4	3	150
Multiple Features (MF)	Handwritten digits	649	10	2000
Image Segmentation (IS)	Outdoor images	19	7	2310
Water Treatment Plant (WTP)	Urban wastewater treatment dataset	38	13	527

(2) Rand Index (RI) [24, 25]

$$\text{RI} = \frac{f_{00} + f_{11}}{N(N-1)/2}, \quad (20)$$

where  $f_{11}$  represents the number of pairing points that have the same class label and belong to the same class and  $f_{00}$  represents the number of matching points with different class labels and belonging to different classes of data points.

The value range of the above two indexes is [0 1]. The closer the value of these two indicators is to 1, the better the performance is. Experimental environment: the experimental hardware platform was Intel Core i7 CPU, with a memory of 16GB. The programming environment is MATLAB 2010.

**4.2. Experimental Results.** In this section, some real-world datasets from the famous UCI database will be used to test our algorithm: (1) Iris dataset, (2) Multiple Features (MF) dataset, (3) Image Segmentation (IS) dataset, and (4) Water Treatment Plant (WTP) dataset. The performance of the Co-MVFCM algorithm proposed in this study is verified and analyzed by using the above datasets when processing real multiview clustering tasks. In order to have a more intuitive impression of the perspectives contained in the three datasets, this paper will present the composition of the four datasets, as shown in Table 1. At the same time, the experimental results of algorithm comparison for these four real datasets are shown in Table 2.

For the Iris dataset, we will observe that the proposed Co-MVFCM has the best clustering performance among all the adopted comparison algorithms. The experimental result of Iris shows that the proposed two multiview collaborative clustering strategies have significant advantages in multiview

TABLE 2: Comparison of performance indices NMI and RI of several algorithms on the datasets.

Dataset	Algorithms	NMI-mean	NMI-std	RI-mean	RI-std
Iris	LSSMTC	0.7260	0.0470	0.8879	0.0219
	CombKM	0.7412	0.0537	0.8636	0.0511
	Coclustering	0.7582	1.17e-016	0.8797	0
	Co-FKM	0.8366	1.48e-016	0.9341	0
	Co-MVFCM	<b>0.8666</b>	<b>0.0005</b>	<b>0.9488</b>	<b>0.0045</b>
MF	LSSMTC	—	—	—	—
	CombKM	0.6883	0.0337	0.8991	0.0119
	Coclustering	0.7024	0.0287	0.9104	0.0126
	Co-FKM	0.7868	0.0715	0.9417	0.0226
	Co-MVFCM	<b>0.7892</b>	<b>0.0217</b>	<b>0.9533</b>	<b>0.0089</b>
IS	LSSMTC	—	—	—	—
	CombKM	0.4815	0.0283	0.7981	0.0176
	Coclustering	0.5182	0.0105	0.8189	0.0208
	Co-FKM	0.5422	0.0353	0.8293	0.0195
	Co-MVFCM	<b>0.5902</b>	<b>0.0217</b>	<b>0.8439</b>	<b>0.0174</b>
WTP	LSSMTC	—	—	—	—
	CombKM	0.1761	0.0121	0.7042	0.0037
	Coclustering	0.2029	0.0103	0.7051	0.0061
	Co-FKM	0.1986	0.0140	0.7056	0.0044
	Co-MVFCM	<b>0.2204</b>	<b>0.0102</b>	<b>0.7209</b>	<b>0.0035</b>

clustering task. For the other three datasets, since the LSSMTC algorithm needs to ensure that the dimensions of each clustering task are consistent, it cannot be used in the face of the samples with different perspectives such as MF, IS, and WTP. By observing the rest of each other algorithm's clustering results of MF datasets, it can be found that based

on the multiple points of view of Co-FKM, the algorithm of this paper has a larger cluster advantage, but because of MF data, no angle exists obvious separability which exist the importance degree of the equilibrium between different points of view; this makes the clustering results from the NMI and RI of the proposed algorithm in the paper with Co-FKM algorithm similar to the average of the two major indicators, and from the variance analysis, the method is still more stable so it still reflects that the method still has certain advantages. For IS dataset, the effect of the proposed method on this sample is relatively obvious, and its clustering index is significantly better than that of the other algorithms, which further confirms the effectiveness of the Co-MVFCM. Finally, through the analysis of the experimental results of the WTP dataset, the same conclusion can be obtained with the above two datasets. In conclusion, through the experiments on real dataset multiple points of view and analysis, we can get a clear conclusion of the clustering algorithm in dealing with multiple points of view which have many view feature clustering tasks generally superior to the clustering algorithm, multiple points of view, and has a view of selective Co-MVFCM algorithm, and clustering algorithm is much better than the previous view.

## 5. Conclusion

Based on cluster technology, multiple points of view are introduced on the basis of the classical FCM algorithm using the Havrda-Charvat entropy structure of different view space approaching. The proposed Co-MVFCM method can better find out the similarities between view compositions, but also from the view of entropy approximation of a different view space, reasonable physical explanation, and thus get more guiding significance to the overall space partition. In addition, in this paper, another contribution is to obtain the importance degree of each view. Through the understanding of the fuzzy theory, multiple points of view are proposed based on the fuzzy index of the adaptive weighted strategy and succeeded in introducing the strategy to the latest fuzzy clustering technology, multiple points of view on the new objective function to achieve the optimal solution. Next, we can evaluate the degree of importance of each view according to the relationship between the weights of each view. The obtained degree of importance of each view provides a new method of integration of the global weighted view space integration means. Experimental results on four real UCI datasets show that the Co-MVFCM has better sample adaptability and superior algorithm performance compared with previous algorithms and related algorithms. However, since the algorithm in this paper is based on the framework of the classical fuzzy *c*-means (FCM) algorithm, the effectiveness of the algorithm may be tested to a certain extent when dealing with higher-dimensional data, which also points out the direction for our future research on the multiview clustering method under high-dimensional data scene.

## Data Availability

The dataset analyzed for this study can be found in this link [<http://archive.ics.uci.edu/ml/index.php>].

## Conflicts of Interest

The authors declare no conflicts of interest.

## Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 61772241 and in part by the 2018 Six Talent Peaks Project in Jiangsu Province under Grant XYDXX-127.

## References

- [1] J. Hämmäläinen, T. Kärkkäinen, and T. Rossi, "Improving scalable K-means<sup>++</sup>," *Algorithms*, vol. 14, no. 1, p. 6, 2021.
- [2] T. Ni, M. Qiao, Z. Chen, S. Zhang, and H. Zhong, "Utility-efficient differentially private K-means clustering based on cluster merging," *Neurocomputing*, vol. 424, pp. 205–214, 2021.
- [3] A. Chen and H. Yan, "An Improved Fuzzy C-means clustering for brain MR images segmentation," *Journal of Medical Imaging and Health Informatics*, vol. 11, no. 2, pp. 386–390, 2021.
- [4] Y. Jiang, K. Zhao, K. Xia et al., "A novel distributed multitask fuzzy clustering algorithm for automatic MR brain image segmentation," *Journal of Medical System*, vol. 43, no. 5, pp. 1–9, 2019.
- [5] X.-B. Zhi, J.-L. Fan, and F. Zhao, "Fuzzy linear discriminant analysis-guided maximum entropy fuzzy clustering algorithm," *Pattern Recognition*, vol. 46, no. 6, pp. 1604–1615, 2013.
- [6] L. Li, H. Ji, and X. Gao, "Maximum entropy fuzzy clustering with application to real-time target tracking," *Signal Processing*, vol. 86, no. 11, pp. 3432–3447, 2006.
- [7] B. Fanyu, C. Hu, Q. Zhang, C. Bai, and L. T. Yang, "Thar Baker: a cloud-edge-aided incremental high-order possibilistic *c*-means algorithm for medical data clustering," *IEEE Transactions on Fuzzy Systems*, vol. 29, no. 1, pp. 148–155, 2021.
- [8] H. Yu, J. Fan, and R. Lan, "Suppressed possibilistic *c*-means clustering algorithm," *Applied Soft Computing*, vol. 80, pp. 845–872, 2019.
- [9] F.-Q. Li, S.-L. Wang, and G.-S. Liu, "A Bayesian possibilistic C-means clustering approach for cervical cancer screening," *Information Sciences*, vol. 501, pp. 495–510, 2019.
- [10] Q. Zhang, L. T. Yang, Z. Chen, and P. Li, "High-order possibilistic *c*-means algorithms based on tensor decompositions for big data in IoT," *Information Fusion*, vol. 39, pp. 72–80, 2018.
- [11] K. D. Koutroumbas, S. D. Xenaki, and A. A. Rontogiannis, "On the convergence of the sparse possibilistic C-means algorithm," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 1, pp. 324–337, 2018.
- [12] H. Wang, H. Shan, and A. Banerjee, "Bayesian cluster ensembles," in *Proceedings of the Ninth SIAM International Conference on Data Mining*, pp. 211–222, John Ascuaga's Nugget, Sparks, Nevada, USA, 2009.
- [13] Y. Jiang, F.-L. Chung, S. Wang, Z. Deng, J. Wang, and P. Qian, "Collaborative fuzzy clustering from multiple weighted views," *IEEE Transactions on Cybernetics*, vol. 45, no. 4, pp. 688–701, 2015.
- [14] Y. Shachor, H. Greenspan, and J. Goldberger, "A mixture of views network with applications to multi-view medical imaging," *Neurocomputing*, vol. 374, pp. 1–9, 2020.



- [15] J. Zhu, K. Li, K. Xia et al., “A novel double-index-constrained, multi-view, fuzzy-clustering algorithm and its application for detecting epilepsy electroencephalogram signals,” *IEEE Access*, vol. 7, pp. 103823–103832, 2019.
- [16] Y. Yamanishi, J. P. Vert, and M. Kanehisa, “Protein network inference from multiple genomic data: a supervised approach,” *Bioinformatics*, vol. 20, Supplement 1, pp. i363–i370, 2004.
- [17] W. Pedrycz, “Collaborative fuzzy clustering,” *Pattern Recognition Letters*, vol. 23, no. 14, pp. 1675–1686, 2002.
- [18] J. Heer, “Mining the structure of user activity using cluster stability,” in *Proc. SIAM Conf. Data Mining, Web Analytics Workshop*, Chicago, Llinois, USA, 2002.
- [19] B. Long, P. S. Yu, and Z. M. Zhang, “A general model for multiple view unsupervised learning,” in *Proc. 8th SIAM Int. Conf. Data Mining*, pp. 822–833, Atlanta, GA, 2008.
- [20] G. Cleuziou, M. Exbrayat, L. Martin, and J.-H. Sublemontier, “CoFKM: a centralized method for multiple-view clustering,” in *Proceedings of the 9th IEEE International Conference on Data Mining (ICDM 2009)*, pp. 752–757, Miami, Florida, USA, 2009.
- [21] J. Chen, S. Chen, Z. Liu, C. Luo, Z. Jing, and Q. Xu, “Health monitoring of landing gear retraction/extension system based on optimized fuzzy C-means algorithm,” *IEEE Access*, vol. 8, pp. 219611–219621, 2020.
- [22] Q. Gu and J. Zhou, “Learning the shared subspace for multi-task clustering and transductive transfer classification,” in *Proceeding of the Ninth IEEE International Conference on Data Mining*, pp. 159–168, Miami Beach, FL, USA, 2009.
- [23] Q. Gu and J. Zhou, “Co-clustering on manifolds,” in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, June 28-July 01*, Paris, France, 2009.
- [24] J. Liu, J. Mohammed, J. Carter, S. Ranka, T. Kahveci, and M. Baudis, “Distance-based clustering of CGH data,” *Bioinformatics*, vol. 22, no. 16, pp. 1971–1978, 2006.
- [25] R. Hathaway, J. Bezdek, and W. Tucker, “An improved convergence theorem for the fuzzy c-means clustering algorithms,” in *Analysis of Fuzzy Information, vol, III*, J. Bezdek, Ed., pp. 123–131, CRC Press, Boca Raton, 1987.

## Research Article

# PPANet: Point-Wise Pyramid Attention Network for Semantic Segmentation

Mohammed A. M. Elhassan <sup>1</sup>, YuXuan Chen,<sup>1</sup> Yunyi Chen,<sup>1</sup> Chenxi Huang <sup>1</sup>, Jane Yang,<sup>2</sup> Xingcong Yao,<sup>1</sup> Chenhui Yang <sup>1</sup> and Yinuo Cheng <sup>3</sup>

<sup>1</sup>School of Informatics, Xiamen University, Xiamen, Fujian Province 361005, China

<sup>2</sup>Department of Cognitive Science, University of California, San Diego, USA

<sup>3</sup>Beijing Jingwei Hirain Technologies Co., Inc, China

Correspondence should be addressed to Chenxi Huang; [supermonkeyxi@xmu.edu.cn](mailto:supermonkeyxi@xmu.edu.cn), Chenhui Yang; [chyang@xmu.edu.cn](mailto:chyang@xmu.edu.cn), and Yinuo Cheng; [yinuo.cheng@hirain.com](mailto:yinuo.cheng@hirain.com)

Received 7 January 2021; Revised 30 January 2021; Accepted 3 April 2021; Published 30 April 2021

Academic Editor: Khin wee Lai

Copyright © 2021 Mohammed A. M. Elhassan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In recent years, convolutional neural networks (CNNs) have been at the centre of the advances and progress of advanced driver assistance systems and autonomous driving. This paper presents a point-wise pyramid attention network, namely, PPANet, which employs an encoder-decoder approach for semantic segmentation. Specifically, the encoder adopts a novel squeeze nonbottleneck module as a base module to extract feature representations, where squeeze and expansion are utilized to obtain high segmentation accuracy. An upsampling module is designed to work as a decoder; its purpose is to recover the lost pixel-wise representations from the encoding part. The middle part consists of two parts point-wise pyramid attention (PPA) module and an attention-like module connected in parallel. The PPA module is proposed to utilize contextual information effectively. Furthermore, we developed a combined loss function from dice loss and binary cross-entropy to improve accuracy and get faster training convergence in KITTI road segmentation. The paper conducted the training and testing experiments on KITTI road segmentation and Camvid datasets, and the evaluation results show that the proposed method proved its effectiveness in road semantic segmentation.

## 1. Introduction

Advanced driver assistance systems (ADAS) have gained massive popularity in the past decades, with much attention given by big car companies such as Tesla, Google, and Uber. ADAS, including adaptive cruise control (ACC), lateral guidance assistance, collision avoidance, traffic sign recognition, and lane change assistance, can be considered crucial factors in developing autonomous driving systems [1–3]. Early studies have developed to detect lanes using mathematical models and traditional computer vision algorithms. For instance, many algorithms have been developed to work on supervised and unsupervised approaches [4–7]. The current paradigm of research has shifted towards nontraditional machine learning methods, namely, deep learning. Deep learning methods have notable performance improvement and have

been the dominant solution for many academia and industry problems because convolutional neural networks (CNNs) extract robust and representative features. The significant improvement in ADAS and autonomous driving field has been driven by deep learning success, particularly deep convolutional neural networks (CNNs).

Road detection is an essential component of many ADAS and autonomous vehicles. There is much active research focusing on performing road detection [8–19] and wide-ranging algorithms of various representations proposed for this regard. Semantic segmentation has been at the centre of this development. There is a significant amount of research using convolution neural network-based segmentation. As region-based representation [20], encoder-decoder networks [21–26] and several supporting approaches along with these networks have been used, while other supporting techniques

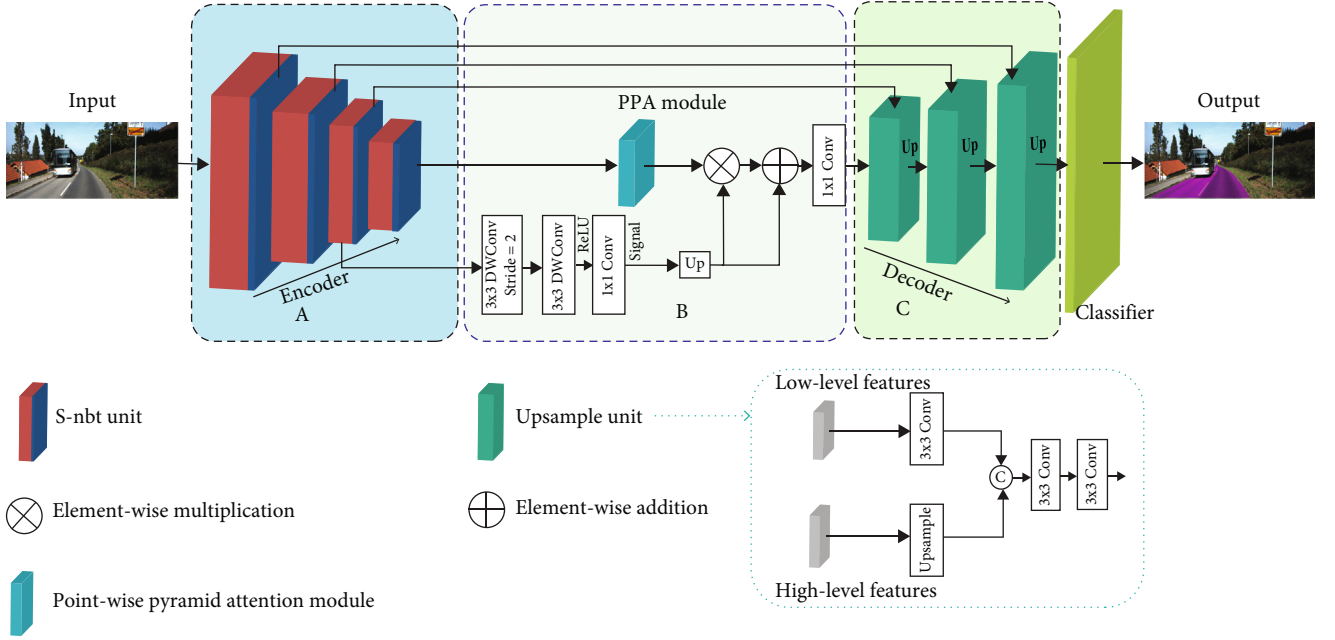


FIGURE 1: Overall architecture of the proposed PPA Net. The encoder adopts squeeze-nbt in an FCN-like network, while PPA and an upsampling unit were employed in the decoder.

fused 3D LiDAR point cloud with 2D images, such as [27, 28]. In this paper, we focus on road segmentation using RGB images. Inspired by the seminal segmentation model U-Net [29], inception [30], SqueezeNet [31], and deep residual learning [32], we propose an architecture that takes the strengths of these well-established models and performs semantic segmentation more effectively. The proposed new architecture is named PPA Net (point-wise pyramid attention network), which follows the encoder-decoder approach. In summary, our main contributions could be summarized as follows.

Firstly, we introduce a novel module named point-wise pyramid attention (PPA module) to acquire long-term dependency and multiscale features without much computation burden. Secondly, we design an upsampling module to help to recover the lost details in the encoder. Thirdly, based on the possibility for improvement, we propose a squeeze-nbt module to extract feature representations in the encoder. At last, we combine these modules in an encoder-decoder manner to construct our PPA Net for semantic segmentation. The designed model was used to improve the performance of road understanding on KITTI road segmentation and Camvid datasets.

## 2. Related Works

**2.1. Encoder-Decoder Method.** In semantic segmentation, the main objective is to assign a categorical label to every pixel in an image, which plays a significant role in road scene understanding. The success and advances in deep convolutional neural network (CNN) models [30, 32–34] have a remarkable impact on pixel-wise semantic segmentation progress due to the rich hierarchical features [29, 35–38]. Usually, to obtain a more delicate result from such a deep network, it

is essential to retain high-level semantic information when using low-level details. However, training such a deep neural network requires a large amount of data, but only a limited number of training examples are available in many practical cases. One way to overcome this problem is by employing transfer learning through a network that is pre-trained on a big dataset then fine-tuned on the targeted dataset, as done in [36, 39]. Another solution for such a problem is performing extensive data augmentation, as done in U-Net [29]. In addition to data augmentation, the model architecture should be designed to propagate the information from low levels to the corresponding high levels in a much easier way, such as U-Net.

**2.2. Deep Neural Networks.** Since the seminal AlexNet [33], model architecture with only eight layers, many studies have been proposed with new approaches for a classification task. Later on, these developed models were applied successfully to a different computer vision task, for example, to segmentation [36], object detection [34], video classification [40, 41], object tracking [42], human pose estimation [43], and super-resolution [44]. These successes spurred the design of a new model with a very large number of layers. However, these growing numbers of layers will need tedious hyperparameter tuning, and that can increase the difficulty of designing such kind of model. In 2014, VGGNet [34] was proposed, in which a significant improvement has been made by utilizing a wider and deeper network; their approach introduced a simple yet effective strategy for designing a very deep network. The quality of a deeper network has a significant impact on improving other computer vision tasks. ResNet [32] has come with an even very deeper model. However, increasing the depth of the network could cause a vanishing gradient problem [32]. Many techniques have been introduced to

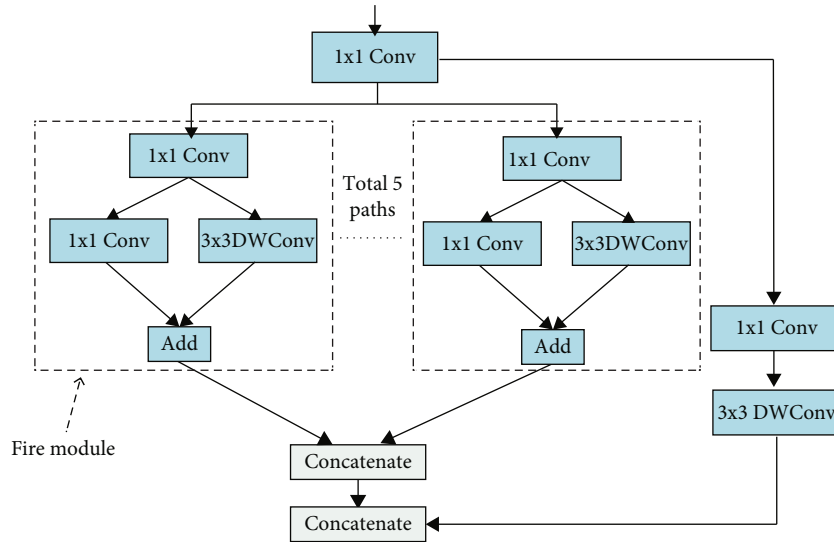


FIGURE 2: Squeeze-nbt module.

prevent vanishing gradients, for instance, using an initialization method MSR [45] and batch normalization [46].

Meanwhile, skip connection (identity mapping) was used to ease the training process of deep networks without vanishing gradient problems, although VGGNet has a simple architecture, which requires high computation capabilities. On the other hand, inception model families have been designed to perform well with constraint memory and low computation budget. In an Inception module, a split transform-merge strategy where the input feature maps are split into lower dimensions (using  $1 \times 1$  convolutions) then transformed by a combination of specialized filters ( $7 \times 7$ ,  $5 \times 5$ , and  $3 \times 3$ ) and merged in the end by concatenating branches is adopted.

**2.3. Semantic Segmentation with CNN.** Recent segmentation-based methods have a significant contribution to solving many computer vision problems, using a wide range of techniques such as a Fully Convolutional Network (FCN) [36], FCN with conditional random field CFD [46], region-based representation [20], encoder-decoder networks [21–23], and multidimensional recurrent networks [47]. Furthermore, pyramid pooling and its variance have a great impact on the recent advances in semantic segmentation [48–53].

**2.4. Dilated Convolution-Based Architecture.** Dilated convolution or atrous convolutions [53] are a powerful tool in the recent progress of semantic segmentation [52]. It is used to enlarge the receptive field while maintaining the same number of parameters. Recently, many approaches focus on multimodal fusion and contextual information aggregation to improve semantic segmentations [52, 54, 55]. ParseNet [56] applies average pooling to the full image to capture the global contextual information. Spatial pyramid pooling (SPP) [57] has inspired the use of pyramid pooling to aggregate multi-scale contextual information such as pyramid pooling [51] module and atrous spatial pyramid pooling module (ASPP) [53, 58]. DenseASPP [59] is proposed to generate dense connections to acquire a larger receptive field. To empower the

TABLE 1: The PPA net network architecture.

Input	Stage	Type	Stride	Output size
				$160 \times 600 \times 3$
Encoder	Stage 1	Squeeze-nbt unit	2	$80 \times 300 \times 32$
		Squeeze-nbt unit	1	$80 \times 300 \times 32$
	Stage 2	Squeeze-nbt unit	2	$40 \times 150 \times 64$
		Squeeze-nbt unit	1	$40 \times 150 \times 64$
	Stage 3	Squeeze-nbt unit	2	$20 \times 75 \times 128$
		Squeeze-nbt unit	1	$20 \times 75 \times 128$
	Stage 4	Squeeze-nbt unit	2	$10 \times 75 \times 256$
		Squeeze-nbt unit	1	$10 \times 75 \times 256$
Decoder	Centre	PPA module		$20 \times 75 \times 128$
	Dec 1	Upsampling unit	—	$20 \times 75 \times 128$
	Dec 2	Upsampling unit	—	$40 \times 150 \times 64$
	Dec 3	Upsampling unit	—	$80 \times 300 \times 32$
	Final	$1 \times 1$ Conv	1	$160 \times 600 \times C$

ASPP module, Xie et al. [60] introduced vortex pooling to utilize contextual information.

### 3. Methodology

**3.1. Architecture.** In this work, we proposed a point-wise pyramid attention network (PPANet) for semantic segmentation, as shown in Figure 1. The network is constructed with an encoder-decoder framework. The encoder is similar to the classification networks; it extracts features and encodes the input data into compact representations. At the same time, the decoder is used to recover the corresponding representations. The squeeze-nbt unit in Figure 2 is used as the main building block for the different stages in the encoder part. Each stage in the encoder has two blocks of the squeeze-nbt unit and the feature map downsampled by half at each first block in each stage using stride convolution

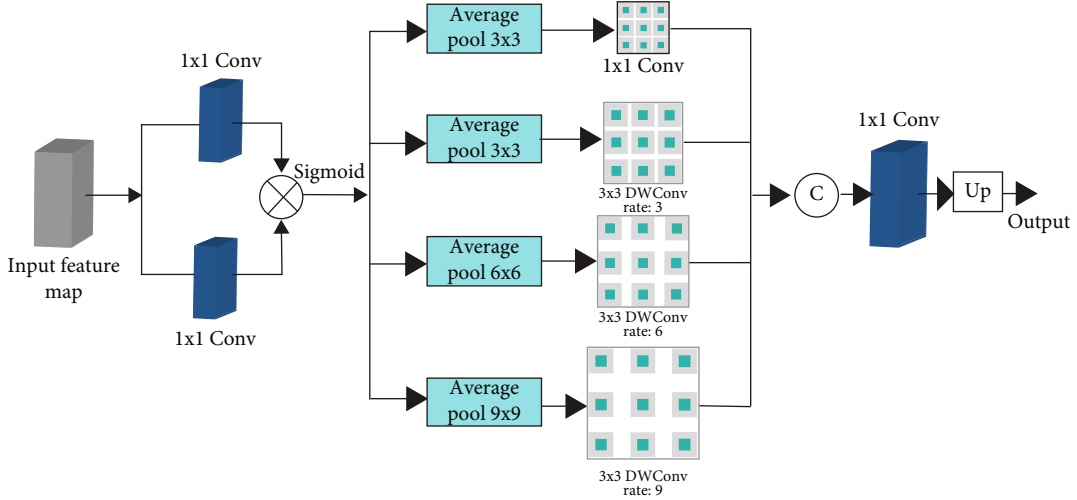


FIGURE 3: Point-wise pyramid attention module.

(for more details, see Table 1. The network has two other parts: the point-wise pyramid attention module and attention module inserted between the encoder and the decoder. These modules in the centre have been used to enrich the receptive field and provide sufficient context information. More details will be discussed in the following sections.

**3.2. Basic Building Unit.** This subsection elaborates the squeeze-nbt module architecture (as illustrated in Figure 2). It draws its inspiration from several concepts that have been introduced into recent state-of-the-art models in classification and segmentation, such as the fire module in SqueezeNet [31], depthwise separable convolution [61], and dilated convolution [58]. Figure 2 is the squeeze-nbt module and encoder-decoder framework. We introduce a new module named squeeze-nbt (squeeze nonbottleneck) module. It is based on a reduce-split-squeeze-merge strategy. The squeeze-nbt module first uses point-wise convolution to reduce the feature maps and then apply a parallel fire module to learn useful representations. To make squeeze-nbt computationally efficient, we adopted  $3 \times 3$  dilated depthwise separable convolution instead of computationally expensive  $3 \times 3$  convolution.

**3.3. Upsampling Module.** Several methods such as [21–23, 62], transpose convolution [63], or bilinear upsampling have been utilized broadly to gradually upsample encoded feature maps. In this work, we proposed the upsample module to work as a decoder and to refine the encoded feature maps by aggregating features of different resolutions. First, the low-level feature is processed with  $3 \times 3$  convolution and in parallel the high-level features upsampled to match the features coming from the encoder; these different features are concatenated and refined with two consecutive  $3 \times 3$  convolutions.

**3.4. Point-Wise Pyramid Attention (PPA) Module.** Segmentation requires both sizeable receptive field and rich spatial

information. We proposed the point-wise pyramid attention (PPA) module, which is effective for aggregating global contextual information. As shown in Figure 3, the PPA module consists of two parts: the nonlocal part and vortex pooling. On the one hand, the nonlocal module will generate dense pixel-wise weight and extract long-range dependency. On the other hand, vortex atrous convolution is useful in detecting an object at multiple scales. By analysing the vortex pooling and nonlocal dependency, we fuse these two modules' advantages in one module named the point-wise pyramid attention (PPA) module. The PPA module consisted of three parallel vortex atrous convolution blocks with dilation rates of 3, 6, and 9 and one nonatrous convolution block.

The point-wise pyramid attention module is shown in Figure 3. Let  $X$  be the input feature map where  $X \in R^{H \times W \times C}$  and  $W$ ,  $H$ , and  $C$  are width, height, and channels, respectively. First, we apply two parallel convolution layers  $F_1 \in R^{H \times W \times C}$  and  $F_2 \in R^{H \times W \times C}$  to generate a feature map of  $F_a \in R^{H \times W \times C'}$ , where  $C' = C/4$  indicates the number of channels of  $F_a$ :

$$\begin{aligned} F_1 &= \text{conv1} \times 1(X), \\ F_2 &= \text{conv1} \times 1(X). \end{aligned} \quad (1)$$

Then, we calculate the similarity matrix  $S \in R^{HW \times HW}$  of  $F_1$  and  $F_1$  by a matrix multiplication  $F_s = F_1 \times F_2^T$ .

Lastly, softmax is applied to normalize the result and transform  $F_s$  to self-attention-like mechanism:

$$\text{Output} = \text{Softmax}(F_s). \quad (2)$$

## 4. Experimental Results and Analysis

In this section, comprehensive experiments on the KITTI road segmentation dataset [64] and Camvid dataset [65] are carried out. We evaluate the efficiency and effectiveness of

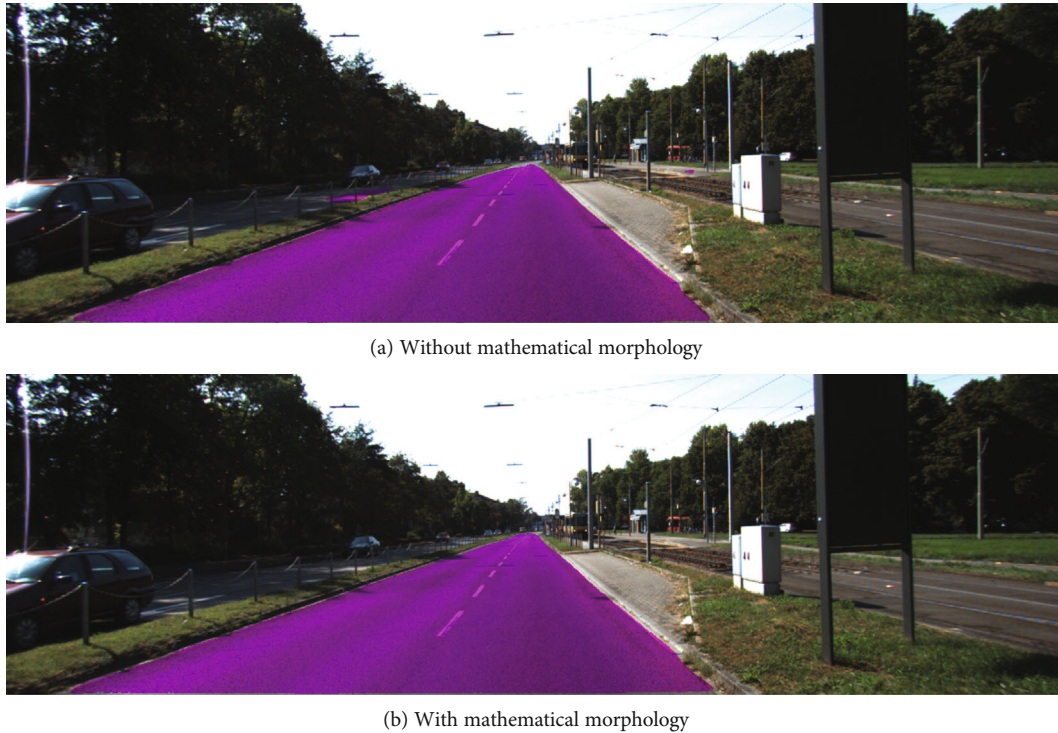


FIGURE 4: Comparison between PPA Net (a) without mathematical morphology and (b) with mathematical morphology. In the first picture, some pixels belong to nonroad classified as a road.

TABLE 2: Ablation study results on the KITTI dataset.

Method	AP (%)	Precision (%)	Recall	max $F$
PPANet-baseline	96.20	97.1	95.6	96.3
PPANet, $r = 3, 6, 13$	91.95	86.44	94.01	89.98
PPANet, $r = 2, 4, 8$	78.90	90.37	88.35	89.36
PPANet, $r = 2$	92.15	92.68	91.41	92.06
Decoder				
PPANet+upsampling unit	91.50	91.89	94.33	93.01
PPANet+PPA	93.54	95.77	95.48	95.16

TABLE 3: Comparison of our model with other networks on the KITTI road segmentation dataset, using average precision (AP), precision, recall, and  $F$ -measure.

Method	AP (%)	Pre (%)	Rec (%)	max $F$
SegNet [23]	89.40	90.90	89.50	90.10
ENet [21]	87.40	88.90	87.60	88.20
FastFCN [68]	95.10	96.10	94.70	95.30
LBN-AA [66]	94.70	96.00	94.10	95.00
DABNet [67]	93.40	94.90	93.50	94.10
AGLNet [69]	94.60	96.00	95.60	96.30
Ours (PPANet)	96.20	97.10	95.60	96.30

our proposed architecture. Firstly, an introduction to the datasets and the implementation protocols is given. We then elaborate on the loss function and the evaluation metrics used to train KITTI, followed by ablation studies and exper-

iments with the SOTA models. Finally, we report a comparison on the Camvid dataset.

#### 4.1. Datasets and Implementation Details

##### 4.1.1. Datasets

(1) *KITTI Road Segmentation Dataset*. It consists of 289 training images with their corresponding ground truth. The data in this benchmark is divided into 3 categories: urban marked (UM) with 95 frames, urban multiple marked lane (UMM) with 96 frames, and urban unmarked (UU). The dataset has a small number of frames and difficult lightning conditions, which make it very challenging. In total, it has 290 frames for testing (testing frames have no ground truth information). Training and testing frames were extracted from the KITTI dataset [64] at a minimum spatial distance of 20 m. Each image has a resolution of  $375 \times 1242$ . We split the dataset into three subsets: (a) training samples with 173 images, (b) validation samples with 58 images, and (c) testing samples with 58 images.

(2) *Camvid Dataset*. The Camvid dataset is an urban street scene understanding dataset in autonomous driving. It consists of 701 samples: 376 training samples, 101 validation samples, and 233 test samples, with 11 semantic categories such as building, road, sky, and bicycle, while class 12 contains unlabelled data that we ignore during training. The original image resolution for the Camvid dataset is  $960 \times 720$ . It has been downsampled into 360x before training for a fair comparison. A weighted categorical cross-entropy loss

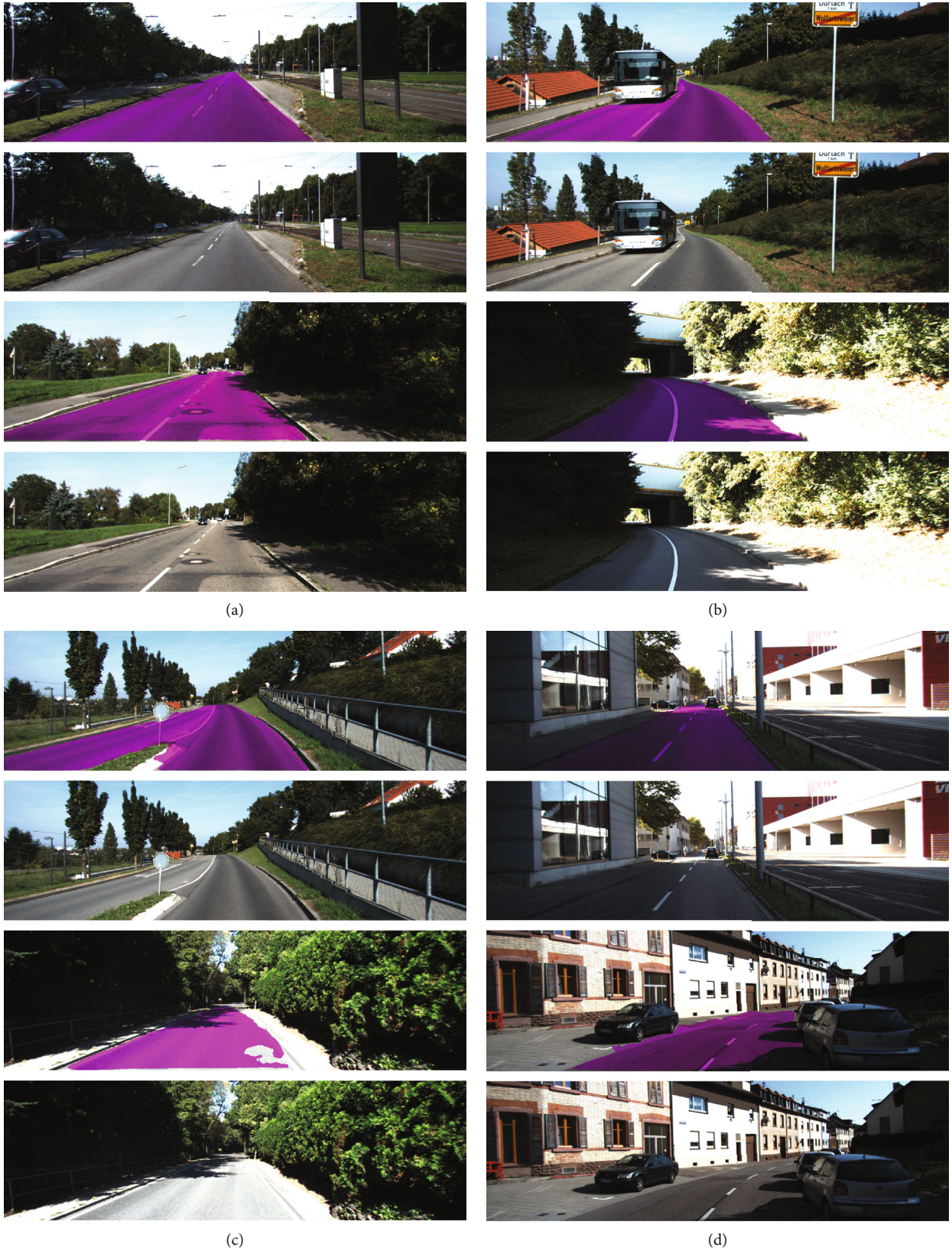


FIGURE 5: Examples of road detection images for the UM test set obtained from the public benchmark suite in perspective view. (a, c) Show the segmentation results; (b, d) show the original images.

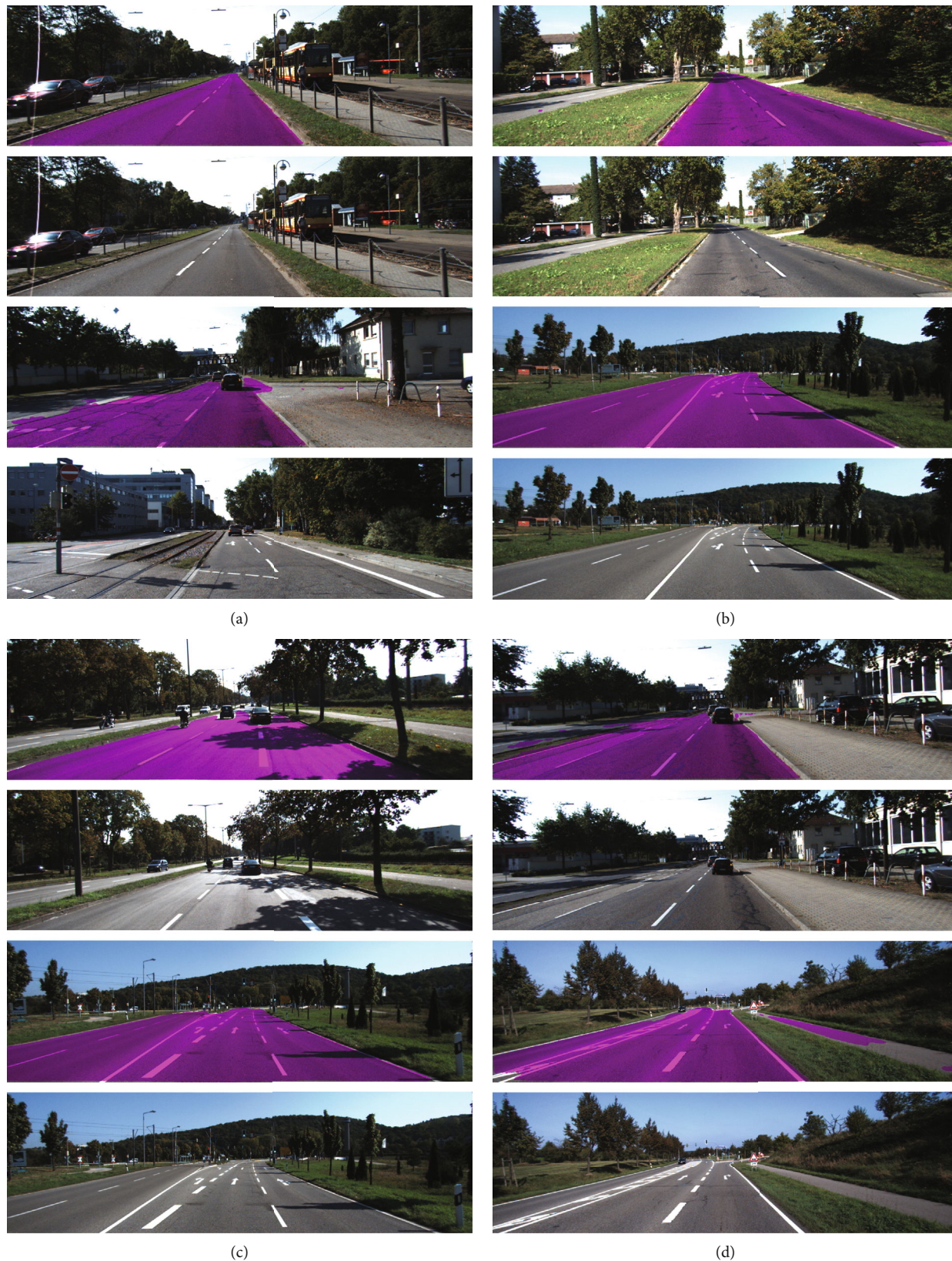


FIGURE 6: Examples of road detection images for the UMM test set obtained from the public benchmark suite in perspective view. (a, c) Show the segmentation results; (b, d) show the original images.



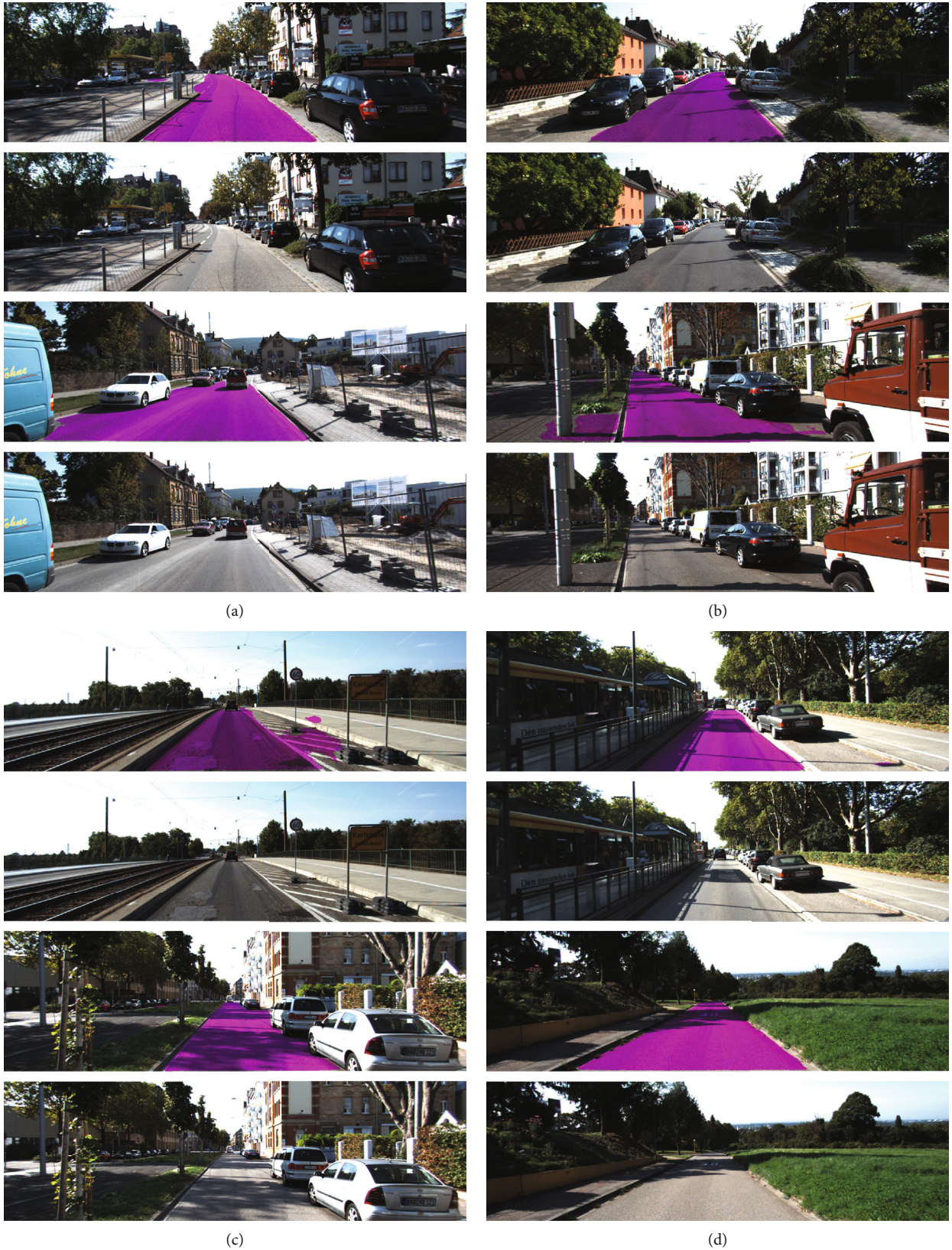


FIGURE 7: Examples of road detection images for the UU test set obtained from the public benchmark suite in perspective view. (a, c) Show the segmentation results; (b, d) show the original images.

TABLE 4: Results of the model on the Camvid dataset.

Method	Year	Params (M)	mIoU (%)
Deeplab-LFOV [58]	2017	262.1	61.6
PSPNet [51]	2017	—	69.1
DenseDecoder [73]	2018	—	70.9
SegNet [23]	2015	29.5	55.6
ENet [21]	2016	0.36	61.3
BiSeNet2 [70]	2018	5.8	68.7
CGNet [71]	2018	0.50	64.0
NDNet45-FCN8-LF [72]	2020	1.1	57.5
LBN-AA [66]	2020	6.2	68.0
DABNet [67]	2019	0.76	66.4
AGLNet [69]	2020	1.12	69.4
PPANet (ours)	—	3.01	70.10

was used to compensate for a small number of categories in the dataset.

**4.1.2. Implementation Details.** All experiments were implemented with one GTX1080Ti CUDA 10.2 and cuDNN 8.0 on Pytorch [58] deep learning framework. The Adam [59] optimizer is a stochastic-based optimizer used with an initial learning rate of  $4e - 4$  to train KITTI road segmentation and Camvid datasets. The learning rate is adjusted according to Equation (3), where  $\alpha$  is the initial learning rate,  $F$  is a factor used to control the learning rate drop,  $D$  is the number of epochs to decrease the learning rate value, and  $i$  is the current epoch. In PPANet implementation, the learning rate is reduced by a factor of every 15 epochs. The proposed network is limited to run for a maximum of 300 epochs. Normal weight initialization [45] is used to initialize the model. Finally,  $7e - 3$  of  $l_2$  regularization to deal with the model overfitting

$$\alpha_{i+1} = \alpha_1 \cdot F^{(1+i)/D}. \quad (3)$$

**(1) Loss Function.** There is a wide range of loss functions proposed over the years to perform semantic segmentation tasks. For instance, binary cross-entropy (BCE) has been applied to many research in classification and segmentation with remarkable success. Although it is convenient to train neural networks using BCE, it might not perform well in class unbalance. For instance, it does not perform well when it is used as the only loss function on KITTI road segmentation with PPANet. In this work, our total loss Equation (6) is a combination of dice loss Equation (5), which was proposed in Zhou et al. [62], and binary cross-entropy Equation (4). Let  $p \in [0, 1]$  be the prediction given by a sigmoid nonlinearity and let  $\hat{p} \in [0, 1]$  be the corresponding ground truth. Dice loss has been implemented in a different form in literature; for instance, in [62, 63], it has equivalent definitions, differing in the denominator value. Our experiment found that using the dice loss function that uses the summation of squared values of probabilities and ground truth in the denominator performs better. These functions are defined as follows.

The binary cross-entropy:

$$\text{BCE}(p, \hat{p}) = \sum_i p \log \hat{p} + (1 - p) \log (1 - \hat{p}). \quad (4)$$

Dice loss:

$$D_i(p, \hat{p}) = \frac{\sum_i p_i \hat{p}_i}{\sum_i (p_i^2 + \hat{p}_i^2)}. \quad (5)$$

Total loss:

$$\text{loss}_{\text{total}} = \text{BCE}(p, \hat{p}) + D_i(p, \hat{p}). \quad (6)$$

**4.1.3. Evaluation Metrics on KITTI.** Precision and recall evaluation metrics can be considered one of the most common metrics for evaluating a binary classification; following the methods used in [64, 66, 67], we evaluated our segmentation model using precision Equation (7), recall Equation (8), and  $F$ -measure Equation (11). The evaluation metrics are listed in the following equations:

$$\text{PRE (precision)} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (7)$$

$$\text{REC (recall)} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (8)$$

$$\text{PFR} = \frac{\text{FP}}{\text{TP} + \text{FP}}, \quad (9)$$

$$\text{FNR} = \frac{\text{FN}}{\text{TP} + \text{FN}}, \quad (10)$$

$$F\text{-measure} = \frac{2 \times \text{PRE} \times \text{REC}}{\text{PRE} + \text{REC}}, \quad (11)$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \quad (12)$$

**4.1.4. KITTI Data Augmentation.** Data augmentation comprises a wide range of techniques used to extend the training samples by applying random perturbations and jitters to the original data. In our model, an online data augmentation approach helps the model learn more robust features and increase the generalizability by preventing the model from seeing the same image twice, as slightly random modification to the input data is performed each time. Therefore, we perform a series of data transformation to deal with typical changes in road images, such as texture and colour changes and illumination. In particular, we implemented normalization, blurring, and changing the illumination. Data augmentation can lend itself naturally in the context of computer vision. For example, we can acquire additional training data from the original KITTI road segmentation images by applying the following transforms:

- (1) Transformation that applied to both image and the ground truth

TABLE 5: Per-class results on the Camvid test set in terms of class mIoU scores.

Methods	Bui	Tree	Sky	Car	Sig	Roa	Ped	Fen	Pol	Side	Bic	mIoU (%)
SegNet [23]	88.8	87.3	92.4	82.1	20.5	97.2	57.1	49.3	27.5	84.4	30.7	55.6
ENet [21]	74.7	77.8	95.1	82.4	51.0	95.1	67.2	51.7	35.4	86.7	34.1	51.3
BiSeNet2 [70]	83.0	75.8	92.0	83.7	46.5	94.6	58.8	53.6	31.9	81.4	54.0	68.7
CGNet [71]	79.8	73.2	90.8	81.3	41.6	95.3	52.9	32.9	28.1	81.9	53.9	64.7
NDNet45-FCN8-LF [72]	85.5	84.6	94.8	82.6	39.2	97.4	60.1	37.3	17.6	86.8	53.7	57.5
LBN-AA [66]	83.2	70.5	92.5	81.7	51.6	93.0	55.6	53.2	36.3	82.1	47.9	68.0
DABNet	81.0	74.1	91.1	81.7	43.0	93.8	56.2	37.2	29.4	78.7	56.5	65.7
AGLNet [69]	82.6	76.1	91.8	87.0	45.3	95.4	61.5	39.5	39.0	83.1	62.7	69.4
PPANet (ours)	84.31	77.84	92.06	86.11	51.17	94.85	62.77	41.86	36.92	82.19	61.12	70.1

- (i) Geometric transformations are used to alter the position of the point, such as translation, scaling, and rotations
- (ii) Mirroring (horizontal flip)
- (2) Transformations that applied to the image only since they affect only pixel values
  - (i) Normalize the input image by standardizing each pixel to be in  $[-1, 1]$  range using Equation (13)
  - (ii) Random brightness adjustment
  - (iii) Gaussian blur
  - (iv) Random noise:

$$\text{input} \times \frac{2}{255} - 1.0 \quad (13)$$

(1) *Mathematical Morphology*. Applying deep learning methods to the segmentation of a road sometimes results in some noise. Nonroad could be classified as a road and vice versa. Several mathematical morphology techniques can be used to remove the noise and improve the performance of the model in the testing time. An opening mathematical morphology process with square structuring elements of  $15 \times 15$  sizes was used. It helped the network eliminate some of the nonroad classified as a road (false positive), as illustrated in Figure 4, where (a) represents the performance of the model without augmentation, and we can see the noise by the side of the road, and (b) shows the effect of removing this false positive when training the model.

## 4.2. Ablation Study

4.2.1. *Encoder*. We carried out some ablation studies to highlight the effectiveness of our proposed model structure. The proposed method baseline achieves 96.3% max  $F$  and 96.2% AP; then, we run experiments with different dilation rate settings. First, we gradually increased the dilation rates 3, 9, and 13 in the encoder at stages 2, 3, and 4, respectively, which result in a decrease of 6.32%  $F$ -measure and 4.25% average

precision. To further examine and verify the effectiveness of our method with a range of dilation rates, we employed another combination dilation rates, 2, 4, and 8, which yield the lowest result in the ablation experiments with 89.36%  $F$ -measure and 78.9%. This sequence of dilation rates has given lower results. It seems that the combination of dilation rates is not effective for the PPANet encoder. Finally, we tested our model using a dilation rate of 2 in the three stages of the encoder and yield the best outcome for our model, as shown in Table 2. Therefore, we set the dilation rate of 2 for all three stages in the encoder.

4.2.2. *Decoder*. We test two settings in the decoding part. First, using the upsampling unit comprises bilinear upsampling and convolution to restore the high-resolution feature from low-resolution features; this approach achieved good results. However, still, there is some information lost during downsampling the feature map in the encoding process. To maintain the highest possible global context feature, we designed a point-wise pyramid attention that is used to increase the model prediction performance with both PPA and upsampling unit; the decoder can aggregate information through a fusion of a multiscale feature. Therefore, it effectively captures local and global context features (see Table 2 for further comparison). It can be seen that the proposed upsample unit and the point-wise pyramid attention (PPA) improve the segmentation-based PPANet and helped achieve superior AP, precision, recall, and max  $F$  score compared to other models.

4.3. *Comparing with the SOTA*. In this subsection, we will present the overall qualitative and quantitative assessment of the trained model. The training and evaluation were conducted using the KITTI road segmentation and Camvid datasets. We then compare the model with a selected SOTA model. Table 3 shows the comparison of PPANet with other SOTA models on the KITTI road segmentation dataset. PPANet is designed for road scene understanding, and it is being trained end-to-end. As previously stated, the dataset has a limited amount of data divided into three categories: urban unmarked (UM road), urban multiple marked lanes (UMM road), and urban unmarked (UU road) as one category to help alongside data augmentation to overcome model overfitting. To rank the best performance among the chosen

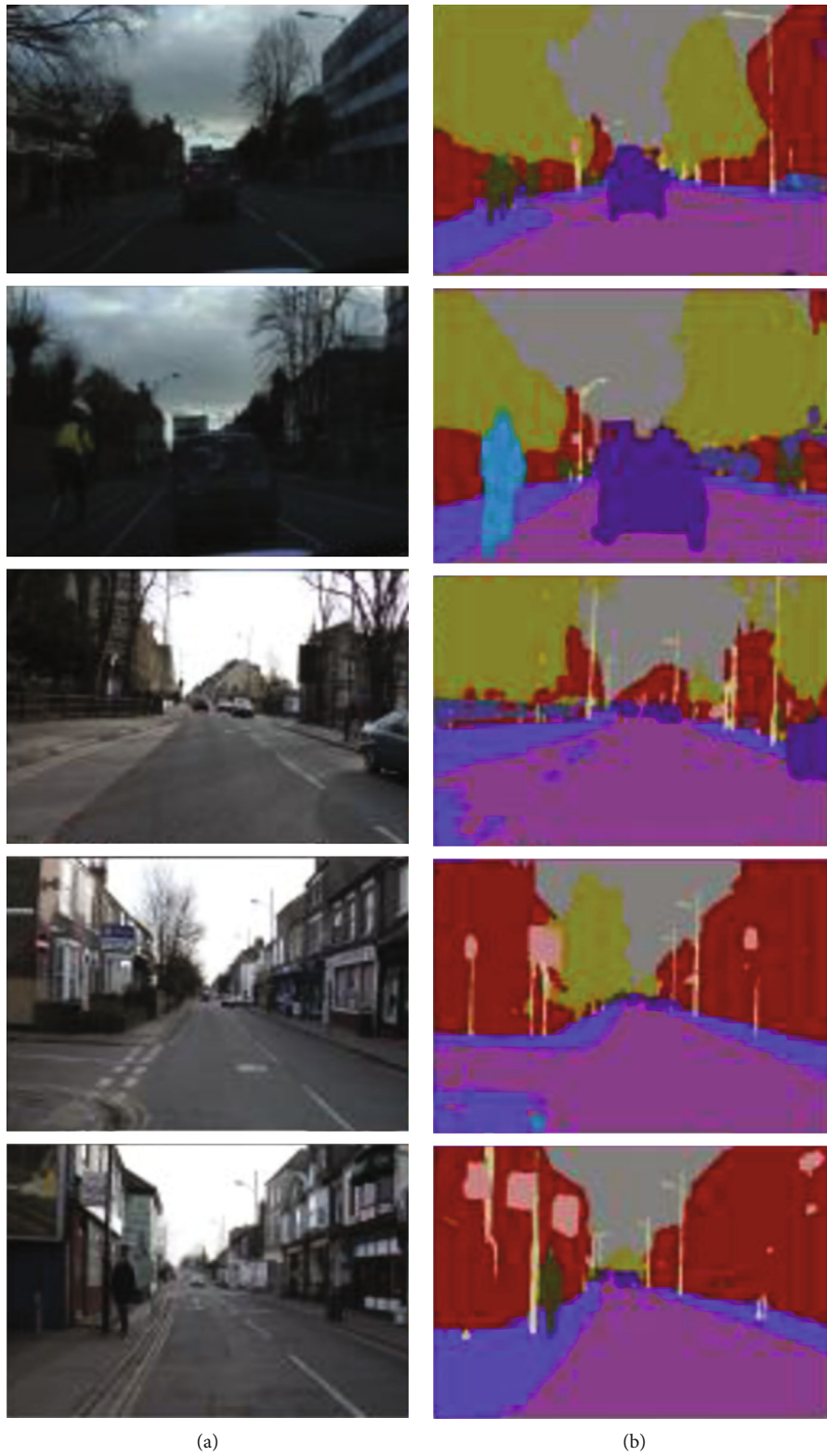


FIGURE 8: Continued.

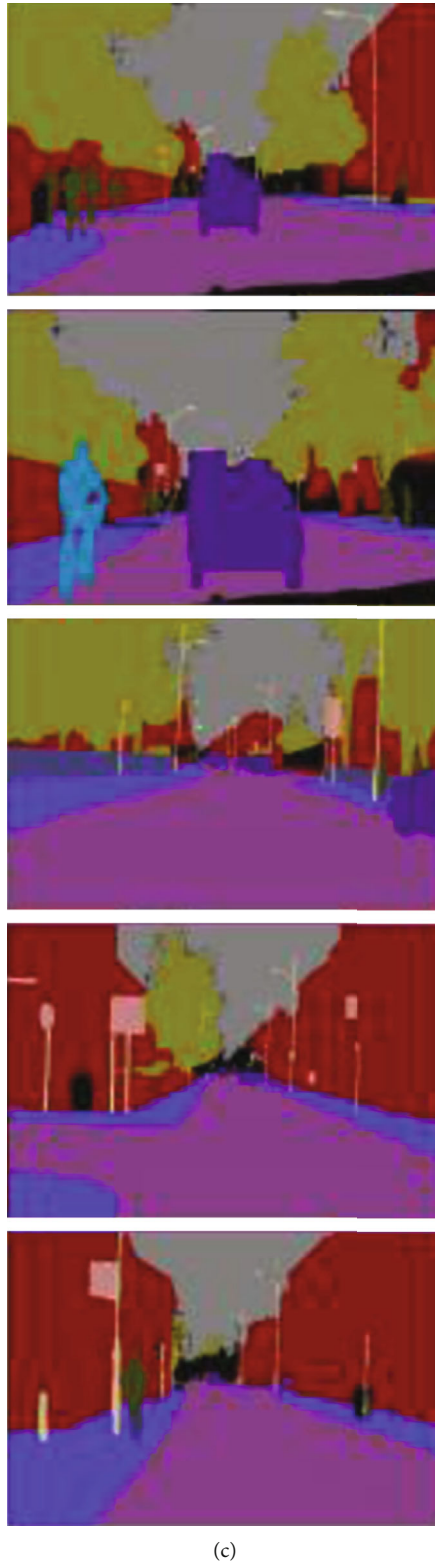


FIGURE 8: Visual results of our method PPANet on the Camvid test set: (a) is the image, (b) is the prediction, and (c) is ground truth.

models for comparison and evaluation, we reported precision (PRE), recall (REC), and max  $F$  metrics, which are known metrics used to evaluate different approaches in binary semantic segmentation. We have chosen some state-of-the-

art models to perform a comparison with our proposed PPANet model. These models include SegNet [23], ENet [21], FastFCN [68], LBN-AA [66], DABNet [67], and AGLNet [69]. The overall results of PPANet and other SOTA models

are illustrated in Figure 3. Our PPA-Net obtained the highest scores for all metrics, demonstrating the effectiveness of the proposed method for robust road detection; FastFCN ranked second in terms of precision and third for max  $F$ , while ALG-Net has better  $f$ -measure. ENet achieved the lowest results compared to all other models. It is designed for speed purposes.

For qualitative performance evaluation of our model in road segmentation, a visual representation of PPA-Net predictions in the KITTI dataset test set is presented in Figures 5–7 in perspective view for UM, UMM, and UU, respectively. We can see that the urban marked (Figure 5) road got the best prediction with almost no misclassification. For urban unmarked, there is little noisy prediction that can be improved using some postprocess optimization techniques such as CRF or increasing the amount of data. When we move to the urban multiple unmarked, it has a higher misclassified road area; it has some area outside the road predicted as road. These false positive detections mainly occur in pole railway when it is close to the road, and also, the road detection is affected by shadow. So, our model with only 3.01 M parameters and without pretrained weights got quite excellent results in a small dataset such as KITTI road segmentation.

**4.4. Comparison with SOTA Models on Camvid.** In this subsection, we design an experiment to demonstrate our proposed network effectiveness and validity on the Camvid dataset. We train and evaluate the model in the training and validation images and validation set for 400 epochs. Then, the model was tested using the testing images and the results reported in Table 4 in terms of mean intersection over union (mIoU). From Table 4, we can see that the proposed PPA-Net method has superior performance in terms of mIoU. First, the model was compared with models that have been designed for real-time semantic segmentation such as ENet [21], BiSeNetv1 [70], CGNet [71], NDNNet45-FCN8-LF [72], LBN-AA [66], DABNet [67], and AGLNet [69]. And also, we compared our proposed method with a non-real-time model such as DeebLabv2 [58], PSPNet [51], DenseDecoder [73], and SegNet [23]. Besides, we present the individual category results in the Camvid test set in Table 5. As can be seen, the proposed method obtained better accuracy in most of the classes. We also provide visual results in Figure 8.

## 5. Conclusion

This paper has presented an approach to scene understanding in monocular images. A novel encoder-decoder network for effective semantic segmentation is proposed, named PPA-Net. The encoder adopts split and squeeze operations in the residual layer to enhance information propagation and feature reuse. To effectively refine the encoded feature map, we design a decoder consisting of the upsampling unit and point-wise pyramid attention (PPA) module. The PPA module is inserted in the centre to enrich the receptive field and to aggregate global contextual information. The attention mechanism is utilized to refine the prediction using a sequence of depthwise convolution followed by sigmoid. This

interaction between different features from the upsampling unit, PPA, and attention provides guidance for high-level and low-level features to improve the performance. The network is trained in an end-to-end manner on two popular datasets: KITTI road segmentation and Camvid. The experimental results showed that the proposed method improves the state of the art for road segmentation on small datasets such as the KITTI dataset and Camvid. Future works will include using pretrained weight as that has been the paradigm for most SOTA in this field. Also, we will investigate the potential of incorporating other sensors such as LiDAR into the architecture and test the effectiveness of our approach in dealing with data fusion and 3D road segmentation.

## Data Availability

We have used the Camvid dataset and KITTI road segmentation dataset.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

The study was funded by the Fujian province Innovation Strategy Research Program (No. 2020R01020196) and Yongtai Artificial Intelligence Institute.

## References

- [1] H. Liu, S. E. Shladover, X.-Y. Lu, and X. Kan, "Freeway vehicle fuel efficiency improvement via cooperative adaptive cruise control," *Journal of Intelligent Transportation Systems*, pp. 1–13, 2020.
- [2] I. Mahdinia, R. Arvin, A. J. Khattak, and A. Ghiasi, "Safety, energy, and emissions impacts of adaptive cruise control and cooperative adaptive cruise control," *Transportation Research Record*, vol. 2674, no. 6, pp. 253–267, 2020.
- [3] Y. Jiang, "Modeling and simulation of adaptive cruise control system," 2020, <https://arxiv.org/abs/2008.02103>.
- [4] E. Kurbatova, "Road detection based on color and geometry characteristics," in *2020 International Conference on Information Technology and Nanotechnology (ITNT)*, pp. 1–5, Samara, Russia, 2020.
- [5] Y. Zhang, L. Wang, H. Wu, X. Geng, D. Yao, and J. Dong, "A clustering method based on fast exemplar finding and its application on brain magnetic resonance images segmentation," *Journal of Medical Imaging and Health Informatics*, vol. 6, no. 5, pp. 1337–1344, 2016.
- [6] Y. Zhang, F.-I. Chung, and S. Wang, "Clustering by transmission learning from data density to label manifold with statistical diffusion," *Knowledge-Based Systems*, vol. 193, article 105330, 2020.
- [7] Y. Zhang, F. Tian, H. Wu et al., "Brain MRI tissue classification based fuzzy clustering with competitive learning," *Journal of Medical Imaging and Health Informatics*, vol. 7, no. 7, pp. 1654–1659, 2017.
- [8] B. Wang, V. Frémond, and S. A. Rodríguez, "Color-based road detection and its evaluation on the KITTI road benchmark," in

- 2014 *IEEE Intelligent Vehicles Symposium Proceedings*, pp. 31–36, Dearborn, MI, USA, 2014.
- [9] L. Geng, J. Sun, Z. Xiao, F. Zhang, and J. Wu, “Combining CNN and MRF for road detection,” *Computers & Electrical Engineering*, vol. 70, pp. 895–903, 2018.
- [10] M. Passani, J. J. Yebeles, and L. M. Bergasa, “CRF-based semantic labeling in miniaturized road scenes,” in *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pp. 1902–1903, Qingdao, China, 2014.
- [11] H. Liu, X. Han, X. Li, Y. Yao, P. Huang, and Z. Tang, “Deep representation learning for road detection through Siamese network,” 2019, <https://arxiv.org/abs/1905.13394>.
- [12] G. L. Oliveira, W. Burgard, and T. Brox, “Efficient deep models for monocular road segmentation,” in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4885–4891, Daejeon, South Korea, 2016.
- [13] F. Ren, X. He, Z. Wei et al., “Fusing appearance and prior cues for road detection,” *Applied Sciences*, vol. 9, no. 5, p. 996, 2019.
- [14] S. Gu, Y. Zhang, X. Yuan, J. Yang, T. Wu, and H. Kong, “Histograms of the normalized inverse depth and line scanning for urban road detection,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 8, pp. 3070–3080, 2018.
- [15] L. Xiao, R. Wang, B. Dai, Y. Fang, D. Liu, and T. Wu, “Hybrid conditional random field based camera-LIDAR fusion for road detection,” *Information Sciences*, vol. 432, pp. 543–558, 2018.
- [16] L. Xiao, B. Dai, D. Liu, D. Zhao, and T. Wu, “Monocular road detection using structured random forest,” *International Journal of Advanced Robotic Systems*, vol. 13, no. 3, p. 101, 2016.
- [17] T. Rateke, K. A. Justen, V. F. Chiarella, A. C. Sobieranski, E. Comunello, and A. V. Wangenheim, “Passive vision region-based road detection,” *ACM Computing Surveys*, vol. 52, no. 2, pp. 1–34, 2019.
- [18] K. Goro and K. Onoguchi, “Road boundary detection using in-vehicle monocular camera,” in *Proceedings of the 7th International Conference on Pattern Recognition Applications and Methods*, pp. 379–387, Funchal, Madeira, Portugal, 2018.
- [19] Y. Lyu, L. Bai, and X. Huang, “Road segmentation using CNN and distributed LSTM,” in *2019 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1–5, Sapporo, Japan, 2019.
- [20] H. Caesar, J. Uijlings, and V. Ferrari, “Region-based semantic segmentation with end-to-end training,” in *European Conference on Computer Vision*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds., vol. 9905 of Lecture Notes in Computer Science, , pp. 381–397, Springer, 2016.
- [21] A. Paszke, A. Chaurasia, S. Kim, and E. Culurciello, “Enet: a deep neural network architecture for real-time semantic segmentation,” 2016, <https://arxiv.org/abs/1606.02147>.
- [22] E. Romera, J. M. Alvarez, L. M. Bergasa, and R. Arroyo, “ERFNet: efficient residual factorized ConvNet for real-time semantic segmentation,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, pp. 263–272, 2018.
- [23] V. Badrinarayanan, A. Kendall, and R. Cipolla, “SegNet: a deep convolutional encoder-decoder architecture for image segmentation,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.
- [24] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, “Unet++: a nested U-net architecture for medical image segmentation,” in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pp. 3–11, Springer, 2018.
- [25] J. Wang, H. Xiong, H. Wang, and X. Nian, “ADSCNet: asymmetric depthwise separable convolution for semantic segmentation in real-time,” *Applied Intelligence*, vol. 50, no. 4, pp. 1045–1056, 2020.
- [26] C. Hazirbas, L. Ma, C. Domokos, and D. Cremers, “Fusenet: incorporating depth into semantic segmentation via fusion-based CNN architecture,” in *Asian conference on computer vision*, S. H. Lai, V. Lepetit, K. Nishino, and Y. Sato, Eds., vol. 10111 of Lecture Notes in Computer Science, , pp. 213–228, Springer, 2016.
- [27] L. Caltagirone, M. Bellone, L. Svensson, and M. Wahde, “LIDAR-camera fusion for road detection using fully convolutional neural networks,” *Robotics and Autonomous Systems*, vol. 111, pp. 125–131, 2019.
- [28] H. Liu, Y. Yao, Z. Sun, X. Li, K. Jia, and Z. Tang, “Road segmentation with image-LiDAR data fusion in deep neural network,” *Multimedia Tools and Applications*, vol. 79, no. 47, pp. 35503–35518, 2020.
- [29] O. Ronneberger, P. Fischer, and T. Brox, “U-net: convolutional networks for biomedical image segmentation,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015*, N. Navab, J. Hornegger, W. Wells, and A. Frangi, Eds., vol. 9351 of Lecture Notes in Computer Science, pp. 234–241, Springer, Cham, 2015.
- [30] C. Szegedy, W. Liu, Y. Jia et al., “Going deeper with convolutions,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9, Boston, MA, USA, 2015.
- [31] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, “SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size,” 2016, <https://arxiv.org/abs/1602.07360>.
- [32] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, Las Vegas, NV, USA, 2016.
- [33] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Advances in neural information processing systems*, vol. 60, no. 6, pp. 84–90, 2017.
- [34] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” 2014, <https://arxiv.org/abs/1409.1556>.
- [35] S. Zheng, S. Jayasumana, B. Romera-Paredes et al., “Conditional random fields as recurrent neural networks,” in *Proceedings of the IEEE international conference on computer vision*, pp. 1529–1537, Santiago, Chile, 2015.
- [36] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431–3440, Boston, MA, USA, 2015.
- [37] Z. Liu, X. Li, P. Luo, C.-C. Loy, and X. Tang, “Semantic image segmentation via deep parsing network,” in *Proceedings of the IEEE international conference on computer vision*, pp. 1377–1385, Santiago, Chile, 2015.
- [38] G. Lin, C. Shen, A. Van Den Hengel, and I. Reid, “Efficient piecewise training of deep structured models for semantic segmentation,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3194–3203, Las Vegas, NV, USA, 2016.
- [39] L. Zhou, C. Zhang, and M. Wu, “D-LinkNet: LinkNet with pretrained encoder and dilated convolution for high resolution

- satellite imagery road extraction,” *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2018, pp. 182–186, Salt Lake City, UT, USA, 2018.
- [40] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, “Large-scale video classification with convolutional neural networks,” in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1725–1732, Columbus, OH, USA, 2014.
- [41] H. Tian, Y. Tao, S. Pouyanfar, S.-C. Chen, and M.-L. Shyu, “Multimodal deep representation learning for video classification,” *World Wide Web*, vol. 22, no. 3, pp. 1325–1341, 2019.
- [42] Y. Wu, J. Lim, and M.-H. Yang, “Online object tracking: a benchmark,” in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2411–2418, Portland, OR, USA, 2013.
- [43] A. Toshev and C. Szegedy, “DeepPose: human pose estimation via deep neural networks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1653–1660, Columbus, Ohio, USA, 2014.
- [44] C. Dong, C. C. Loy, K. He, and X. Tang, “Learning a deep convolutional network for image super-resolution,” in *Computer Vision – ECCV 2014. ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds., vol. 8692 of Lecture Notes in Computer Science, pp. 184–199, Springer, Cham, 2014.
- [45] K. He, X. Zhang, S. Ren, and J. Sun, “Delving deep into rectifiers: surpassing human-level performance on ImageNet classification,” in *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 1026–1034, Santiago, Chile, 2015.
- [46] S. Ioffe and C. Szegedy, “Batch normalization: accelerating deep network training by reducing internal covariate shift,” in *Proceedings of the 32nd International Conference on Machine Learning*, Lille, France, 2015.
- [47] A. Graves, S. Fernández, and J. Schmidhuber, “Multi-dimensional recurrent neural networks,” in *Artificial Neural Networks – ICANN 2007. ICANN 2007*, vol. 4668 of Lecture Notes in Computer Science, pp. 549–558, Springer, Berlin, Heidelberg, 2007.
- [48] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, “Encoder-decoder with atrous separable convolution for semantic image segmentation,” in *Computer Vision – ECCV 2018. ECCV 2018*, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds., vol. 11211 of Lecture Notes in Computer Science, pp. 801–818, Springer, Cham, 2018.
- [49] S. Mehta, M. Rastegari, A. Caspi, L. Shapiro, and H. Hajishirzi, “ESPNet: efficient spatial pyramid of dilated convolutions for semantic segmentation,” in *Computer Vision – ECCV 2018. ECCV 2018*, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds., vol. 11214 of Lecture Notes in Computer Science, pp. 552–568, Springer, Cham, 2018.
- [50] S. Mehta, M. Rastegari, L. Shapiro, and H. Hajishirzi, “ESP-Netv2: a light-weight, power efficient, and general purpose convolutional neural network,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 9190–9200, Long Beach, CA, 2019.
- [51] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, “Pyramid scene parsing network,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2881–2890, Honolulu, HI, USA, 2017.
- [52] F. Yu and V. Koltun, “Multi-scale context aggregation by dilated convolutions,” 2015, <https://arxiv.org/abs/1511.07122>.
- [53] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, “Rethinking atrous convolution for semantic image segmentation,” 2017, <https://arxiv.org/abs/1706.05587>.
- [54] X. Lian, Y. Pang, J. Han, and J. Pan, “Cascaded hierarchical atrous spatial pyramid pooling module for semantic segmentation,” *Pattern Recognition*, vol. 110, article 107622, 2021.
- [55] Y. Zhang, S. Wang, K. Xia, Y. Jiang, P. Qian, and For the Alzheimer’s Disease Neuroimaging Initiative, “Alzheimer’s disease multiclass diagnosis via multimodal neuroimaging embedding feature selection and fusion,” *Information Fusion*, vol. 66, pp. 170–183, 2021.
- [56] W. Liu, A. Rabinovich, and A. C. Berg, “ParseNet: looking wider to see better,” 2015, <https://arxiv.org/abs/1506.04579>.
- [57] K. He, X. Zhang, S. Ren, and J. Sun, “Spatial pyramid pooling in deep convolutional networks for visual recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1904–1916, 2015.
- [58] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, “DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834–848, 2018.
- [59] M. Yang, K. Yu, C. Zhang, Z. Li, and K. Yang, “DenseASPP for semantic segmentation in street scenes,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3684–3692, Salt Lake City, UT, USA, 2018.
- [60] C.-W. Xie, H.-Y. Zhou, and J. Wu, “Vortex pooling: improving context representation in semantic segmentation,” 2018, <https://arxiv.org/abs/1804.06242>.
- [61] F. Chollet, “Xception: deep learning with depthwise separable convolutions,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1251–1258, Honolulu, HI, USA, 2017.
- [62] Q. Zhou, W. Yang, G. Gao et al., “Multi-scale deep context convolutional neural networks for semantic segmentation,” *World Wide Web*, vol. 22, no. 2, pp. 555–570, 2019.
- [63] H. Noh, S. Hong, and B. Han, “Learning deconvolution network for semantic segmentation,” in *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 1520–1528, Santiago, Chile, 2015.
- [64] J. Fritsch, T. Kuehnl, and A. Geiger, “A new performance measure and evaluation benchmark for road detection algorithms,” in *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*, pp. 1693–1700, The Hague, Netherlands, 2013.
- [65] G. J. Brostow, J. Fauqueur, and R. Cipolla, “Semantic object classes in video: a high-definition ground truth database,” *Pattern Recognition Letters*, vol. 30, no. 2, pp. 88–97, 2009.
- [66] G. Dong, Y. Yan, C. Shen, and H. Wang, “Real-time high-performance semantic image segmentation of urban street scenes,” *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–17, 2020.
- [67] G. Li, I. Yun, J. Kim, and J. Kim, “DABNet: depth-wise asymmetric bottleneck for real-time semantic segmentation,” 2019, <https://arxiv.org/abs/1907.11357>.
- [68] H. Wu, J. Zhang, K. Huang, K. Liang, and Y. Yu, “FastFCN: rethinking dilated convolution in the backbone for semantic segmentation,” 2019, <https://arxiv.org/abs/1903.11816>.
- [69] Q. Zhou, Y. Wang, Y. Fan et al., “AGLNet: towards real-time semantic segmentation of self-driving images via attention-



- guided lightweight network,” *Applied Soft Computing*, vol. 96, p. 106682, 2020.
- [70] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, and N. Sang, “BiSeNet: bilateral segmentation network for real-time semantic segmentation,” in *Computer Vision – ECCV 2018. ECCV 2018*, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds., vol. 11217 of Lecture Notes in Computer Science, pp. 325–341, Springer, Cham, 2018.
- [71] T. Wu, S. Tang, R. Zhang, and Y. Zhang, “CGNet: a lightweight context guided network for semantic segmentation,” 2018, <https://arxiv.org/abs/1811.08201>.
- [72] Z. Yang, H. Yu, M. Feng et al., “Small object augmentation of urban scenes for real-time semantic segmentation,” *IEEE Transactions on Image Processing*, vol. 29, pp. 5175–5190, 2020.
- [73] P. Bilinski and V. Prisacariu, “Dense decoder shortcut connections for single-pass semantic segmentation,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6596–6605, Salt Lake City, UT, USA, 2018.

## Research Article

# A Transfer Deep Generative Adversarial Network Model to Synthetic Brain CT Generation from MR Images

Yi Gu  and Qiankun Zheng 

School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi, Jiangsu 214122, China

Correspondence should be addressed to Yi Gu; 8202101437@jiangnan.edu.cn

Received 29 March 2021; Revised 14 April 2021; Accepted 15 April 2021; Published 26 April 2021

Academic Editor: Shan Zhong

Copyright © 2021 Yi Gu and Qiankun Zheng. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Background.** The generation of medical images is to convert the existing medical images into one or more required medical images to reduce the time required for sample diagnosis and the radiation to the human body from multiple medical images taken. Therefore, the research on the generation of medical images has important clinical significance. At present, there are many methods in this field. For example, in the image generation process based on the fuzzy C-means (FCM) clustering method, due to the unique clustering idea of FCM, the images generated by this method are uncertain of the attribution of certain organizations. This will cause the details of the image to be unclear, and the resulting image quality is not high. With the development of the generative adversarial network (GAN) model, many improved methods based on the deep GAN model were born. Pix2Pix is a GAN model based on UNet. The core idea of this method is to use paired two types of medical images for deep neural network fitting, thereby generating high-quality images. The disadvantage is that the requirements for data are very strict, and the two types of medical images must be paired one by one. DualGAN model is a network model based on transfer learning. The model cuts the 3D image into multiple 2D slices, simulates each slice, and merges the generated results. The disadvantage is that every time an image is generated, bar-shaped “shadows” will be generated in the three-dimensional image. **Method/Material.** To solve the above problems and ensure the quality of image generation, this paper proposes a Dual3D&PatchGAN model based on transfer learning. Since Dual3D&PatchGAN is set based on transfer learning, there is no need for one-to-one paired data sets, only two types of medical image data sets are needed, which has important practical significance for applications. This model can eliminate the bar-shaped “shadows” produced by DualGAN’s generated images and can also perform two-way conversion of the two types of images. **Results.** From the multiple evaluation indicators of the experimental results, it can be analyzed that Dual3D&PatchGAN is more suitable for the generation of medical images than other models, and its generation effect is better.

## 1. Introduction

The combination of artificial intelligence and various fields has brought a lot of convenience to mankind. In the medical field, common machine learning algorithms such as decision trees, K-nearest neighbors, and support vector machines [1–5] can assist doctors in the diagnosis and prevention of diseases by processing text data, making the diagnosis and prevention of patients’ diseases more efficient and accurate. However, under medical image data sets such as magnetic resonance imaging (MRI), the implementation effect of the above algorithm is not ideal. Then, deep learning algorithms

were applied to medical image data. The deep learning neural network simulates the process of human brain learning and cognition of external things by analyzing and learning a large number of original images, continuously extracting and discovering abstract high-level features from low-dimensional features. This algorithm can process medical image data sets relatively accurately. However, whether it is traditional machine learning algorithms or the recently emerging deep learning algorithms, they all require a large amount of artificially labeled data as a prerequisite, and they need enough data to provide a guarantee for the accurate results of the algorithm. In the field of medical imaging, due to the low

prevalence of rare diseases and the privacy of patients involved in medical data, it becomes very difficult to obtain medical image sample information. Therefore, the sample diversity of medical image data sets cannot be guaranteed. As we all know, training a good machine learning algorithm requires a sufficient amount of medical image data. However, the difficulty of finding diseased samples and insufficient sample diversity will lead to unbalanced classification and worsen the final classification performance. The methods to solve the problem of image scarcity mainly include traditional image data enhancement, Variational Auto Encoders (VAE) [6, 7], and Generative Adversarial Networks (GANs) [8–12]. Traditional image data enhancement methods can increase the number of image samples to a certain extent, but a large-scale generation of image samples will increase the risk of overfitting. Although the VAE method solves the overfitting problem caused by a single generated image, the VAE method generates blurred images and cannot be used for the next step of medical image research. The reason is that VAE judges the quality of the generated image by directly calculating the mean square error between the generated image and the original image. GANs generate clear and usable medical images through continuous adversarial learning between the generator and the discriminator.

Although traditional image data enhancement techniques can alleviate the problem of the small number of medical samples, they produce highly relevant image training data. Although the VAE method can solve the problem of overfitting caused by a single generated image, it produces blurred images, which affects the usability of medical images. Therefore, whether to find a way to increase the diversity of medical image data and solve the problem of scarcity of medical image data is a problem that needs to be solved urgently. Traditional image data enhancement algorithms have been seldom used in recent years, because traditional image data enhancement methods, such as flip, translation, rotation, and cropping, can only simply increase the size of the data set and cannot add some new image information to the data set. This method does not contribute much to the diversity of image samples and cannot effectively improve the entire medical image algorithm research. Traditional image data enhancement methods can increase the size of image samples to a certain extent, but they cannot fundamentally solve the problem of scarcity of medical image data. Excessive use of traditional image data enhancement methods will also increase the risk of overfitting. Therefore, a large number of researchers began to apply GANs to the field of medical image generation.

In 2016, reference [13] proposed a generative adversarial network model that uses a fully convolutional network as a generator. This model is used in this paper to realize the conversion between MRI images and CT images of brain tumors [14]. References [15] proposed a new depth-differentiated generative adversarial network model for the imbalance of skin lesion image categories-DDGANs. This model integrates the generative countermeasure network structures DCGANs [16] and LAPGANs [17] and uses a single noise source to distinguish the input of multiple noise sources from LAPGANs. References [18] proposed a new pipeline model

based on a generative confrontation network in response to the problem that the current medical image data set is not easy to obtain. The model can generate public and extensive data sets and will not be affected by privacy issues. References [19] proposed a new model Stain GANs based on generative adversarial networks. This model not only eliminates the dependence on reference images but also achieves a high degree of visual similarity with the target domain. Stain GANs is an unpaired image-to-image conversion model based on Cycle GANs [20]. The classification effect of breast tumors proves that the model can perform better conversion between images. This once again proves that the generative confrontation network can be well applied in the field of medical imaging.

The research on the application of GANs in the field of medical imaging has not ended but constantly ushered in new peaks. In 2019, DARSAlman UH [21] and others proposed a multicontrast MRI generation method based on a conditional generative confrontation network. Unlike most traditional methods, this method performs end-to-end training of GANs, which synthesizes a target image with a similar contrast given the contrast of the source image. The use of the adversarial loss function improves the synthesis accuracy of high spatial frequency information in the target contrast. This method uses the information of adjacent cross-sections in each volume to improve the accuracy of the synthesis. GAN-based methods offer great hope for multi-contrast synthesis in clinical practice. GAN is mainly used to increase the number of small samples and unsupervised learning. As more and more generative confrontation networks based on GANs are proposed, we can find that applying GANs to the medical image field can solve the problem of small medical image data. However, the generator structure of GANs is only composed of simple convolutional networks, which cannot quickly generate clear images. And because the original data is scarce, it cannot provide enough samples for the discriminator to train. In response to the above problems, this paper proposes a method based on transfer learning Dual3D&PatchGAN. The difference between this method and other methods is as follows.

- (1) The core idea of the method is different. Neither the fuzzy clustering method [22–26] nor the classic GAN confrontation generation idea uses the transfer learning mechanism [27–30]. Most of them are generated by “point-to-point” algorithms. In practical applications, the scope of such applications is very limited
- (2) The network model is different. In many medical image generation models, a 2-dimensional model is often used for image generation. That is, the image is first cut into slices, a 2D network model is used to generate a simulated image, and the generated simulated slices are synthesized into a 3D image. The Dual3D&PatchGAN model is aimed at a 3D model, and the result generated each time is a three-dimensional block, which is often more accurate for the model result. In addition, the implementation

details are different. In the module that generates the network, when calculating the Softmax value, the entire image is often calculated. In this model, the original image is divided into several regions, and the Softmax value of each domain is calculated to obtain the average value

- (3) The requirements of the data set are different. The traditional GAN model often requires paired training, that is, two medical images at the same location are required, and the shooting position angles should be strictly the same. The requirements for the data set are almost harsh. The Dual3D&PatchGAN model only needs two types of medical images and does not have too many requirements, which is meaningful for clinical practice
- (4) The conversion efficiency is different. The traditional GAN model can often convert CT images into MR images, which is a one-way process. When the model is trained each time, only one-way image conversion can be performed. The Dual3D&PatchGAN model is a conversion between two domains, which can be converted in both directions. In terms of conversion efficiency, the Dual3D&PatchGAN model is more efficient

## 2. Backgrounds

In 2014, Goodfellow proposed the GANs model. The network is mainly composed of two core parts, one is the generative model  $G$  used to generate fake samples, and the other is the discriminant model  $D$  used to determine the authenticity. The structure of the network is shown in Figure 1.

The generator and discriminator in the GANs model compete with each other, and the generator generates a sample  $x = G(z; \theta_g)$  based on the source picture. As a competitor, the discriminator  $p = D(z; \theta_d)$  must generate samples for the discrimination and distinguish whether the generated samples are real or fake. The output of the discriminator is a probability value in the range of  $[0, 1]$ . The probability value is used to indicate the possibility that the input sample in the discriminator is a real sample. The learning process of GANs is a process of adversarial learning. The loss function of the discriminator is represented by  $-V(D, G)$ , and the loss function of the generator is represented by  $V(D, G)$ . In the GANs learning process, both the generator and the discriminator are trying to minimize their respective losses. The optimal solution of the entire network is defined as follows:

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log (1 - D(G(z)))]. \quad (1)$$

During the training process, the generator continuously learns the ability to generate images, trying to make the discriminator judge the image generated by itself as a real sample, so that  $D(G(z))$  approaches 1. However, the training of the discriminator is a binary classification problem, which

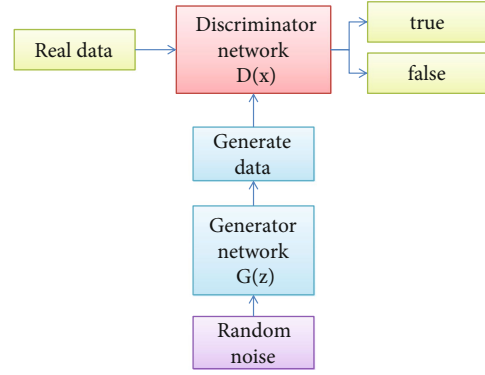


FIGURE 1: GANs network structure diagram.

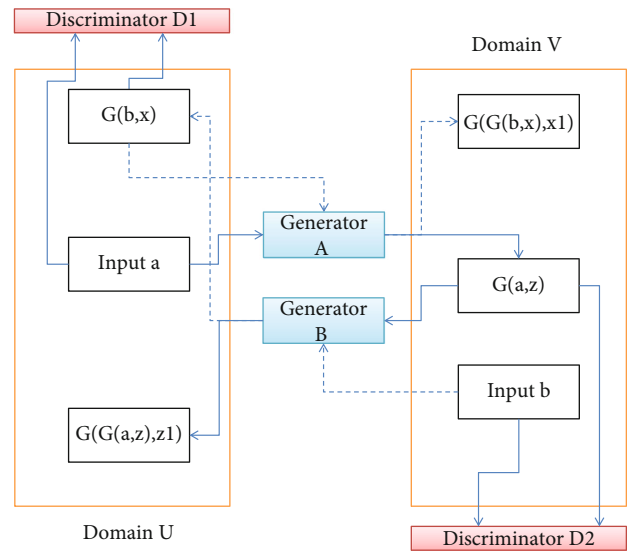


FIGURE 2: Dual3D&PatchGAN model structure diagram.

attempts to clearly distinguish between real data and generated data. It is hoped that the output  $D(x)$  of the real data will be close to 1, and the output  $D(G(z))$  of the generated data will be close to 0. When the output of the discriminator for each input sample is  $1/2$ , it can no longer distinguish between fake samples and real samples. At this point, the entire network model training reaches convergence, and the model training ends.

## 3. Dual3D&PatchGAN Model

The images generated by the DualGAN model [31–33] based on transfer learning do not need to consider that the data sets must be paired one by one. At the same time, the quality of generating simulated medical images is higher. But as far as the experimental effect is concerned, the image generated by this method is easy to cause horizontal stripes. This article improves DualGAN so that its convolution is no longer limited to the two-dimensional model but focuses on the three-dimensional space. At the same time, improve its discrimination network and add PatchGAN module to further improve network performance. This paper proposes a

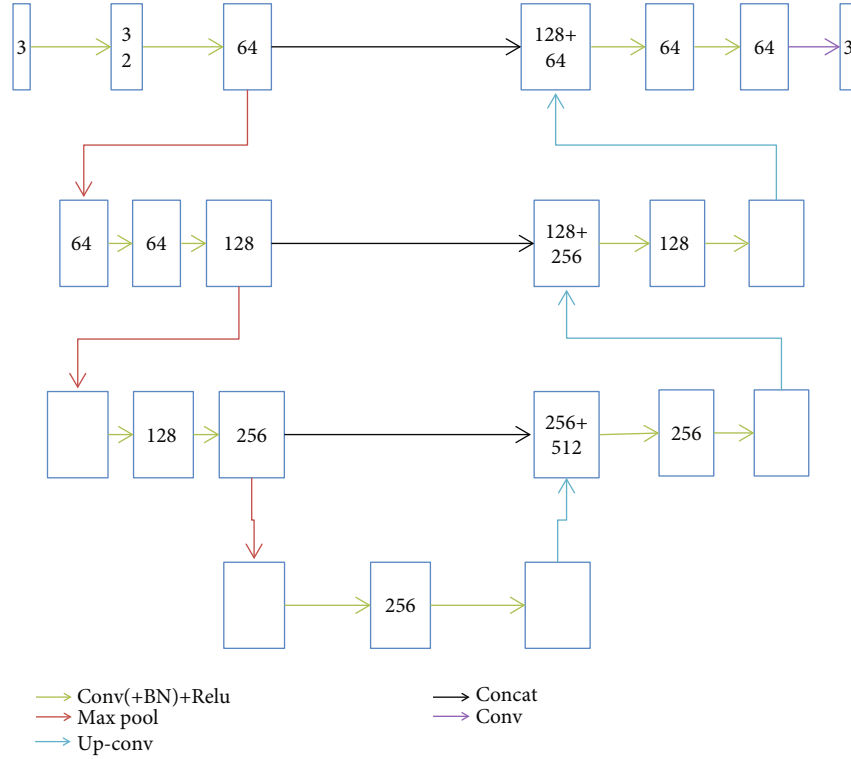


FIGURE 3: Dual3D&amp;PatchGAN generative network (G) model.

TABLE 1: Dual3D&amp;PatchGAN generation network (G) parameters.

AH	Value	SH	Value
Convolution block	$3 \times 3 \times 3$	Convolution block	$3 \times 3 \times 3$
Downsampling	$2 \times 2 \times 2$ max pooling	Up-sampling	$2 \times 2 \times 2$ max pooling
Step size	2	Step size	2
Activation function	Rectified linear unit, ReLU	Activation function	Rectified linear unit, ReLU

network model Dual3D&PatchGAN based on 3DGAN network and transfer learning. The model structure is shown in Figure 2. The network is composed of two pairs of generating network (G) and discriminating network (D). The task of generating network (G) is to continuously generate fake data to deceive and discriminate the network as much as possible, while the task of discriminating network (D) is to discriminate the fake images generated by the generating network as much as possible. At the same time, the two pairs of generating networks and discriminating networks also enable the two types of images to be converted to each other at the same time. In summary, the image conversion process of the Dual3D&PatchGAN model is carried out in both directions at the same time, and its 3D generation network (G) is shown in Figure 3. The goal of this model training is to make the two networks fit as closely as possible, and there will be multiple loss functions to limit the network model. Part of the loss function is shown in Eqs. (2) and (3).

$$\text{Loss}(\text{before}) = \|b - G(G(b, x), x1)\|, \quad (2)$$

$$\text{Loss}(\text{after}) = \|a - G(G(a, z), z1)\|. \quad (3)$$

To illustrate the relevant parameters of the model, Table 1 lists the parameters of the generating network (G) in detail. The generative network (G) is divided into the first half and the second half, the first half is the analysis path (AH), and the second half is the synthesis path (SH).

The experimental process of Dual3D&PatchGAN model is shown in Figure 4. The implementation steps of the model proposed in this article are as follows:

- (1) *Image Preprocessing*. First, determine the medical images of the two domains. In this experiment, the medical CT domain and the medical MR domain are used. Second, collect images of the selected domain. When collecting, pay attention to selecting several images with clear details, consistent size, and excellent quality for the training set. At the same time, some images should be prepared for the test set. Since there are 9 sample data in the data set used, 8 sample data is used as the training set and 1 sample

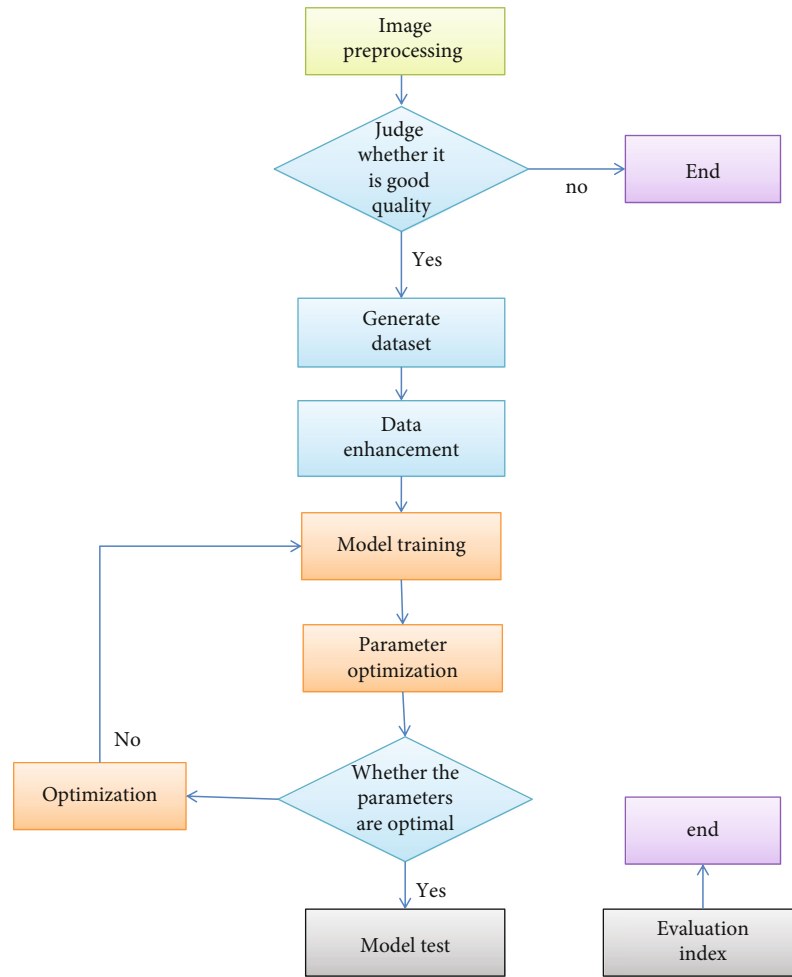


FIGURE 4: Experimental flowchart of Dual3D&amp;PatchGAN model.

data is used as the test set each time, and the ratio of the training set to the test set is 8:1

- (2) *Make a Data Set.* Filter the images in the two selected domains again to remove blurry, artifacts and other unclear images in the image. The two types of medical images are placed in two folders, and the two types of images are marked with domains for model training. Type conversion of the original image. The original image format is .png, which needs to be converted to a .gif sequence. This is because the Dual3D&PatchGAN model is a 3-dimensional model, and the 3-dimensional space must be convolved
- (3) *Data Enhancement.* The original data set is expanded by means such as rotation and reflex. In this way, the model can learn as much knowledge as possible and increase the generalization ability of the model
- (4) *Model Training.* The images in the two domains are used as the two types of image input of the model, and 800 rounds of model training are performed
- (5) *Observation Model.* Since the entire model is created using the Tensorflow framework, Tensorboard can be used to observe the details of the entire model training and gradually determine the range of model parameters
- (6) *Grid Optimization.* For a specific data set, the network will have different degrees of deviation, so grid optimization is required. Its main function is to help the network find the best parameters for a specific data set. Through the above steps, the optimal value range of the network parameter has been determined, and the optimal value of the model parameter is gradually determined in this range
- (7) *Repeat the Training.* The grid optimization method is used to find the most suitable parameters, which are used for model training
- (8) *Check Again.* Observe the training details of the model again through Tensorboard to check whether the network model has reached the best fit. If it has not reached the best fit, readjust the model parameters

TABLE 2: Experimental environment.

Experimental environment	Parameter	Experimental environment	Parameter
Development system	Windows, Linux (Ubuntu 16.04)	CPU	Intel i7-6850K 3.60 GHz
Development environment	Matlab2018b, Pycharm	GPU	Ubuntu 16.04 (64 bit), 128 GB of RAM, NVIDIA TITAN XP(12 GB) GPU card
Development language	Python, matlab	Training time	About 31.6 hours
Model initial learning rate	1E-5	Training set	8 samples for each of the two types
Compilation framework	Tensorflow, Keras	Test set	1 sample for each of the two types
Training times	800 times	Network model	Dual3D&PatchGAN
Model dimensions	3D		

(9) *Data Testing*. The test set tests the trained model and visually sees the pros and cons of the test results

(10) *Evaluation Indicators*. The test result of the test set obtained through the above steps is the simulated image generated by the model. The evaluation index calculation formula is used to calculate multiple evaluation indexes. The evaluation indicators used in this experiment are structural similarity (SSIM) and peak signal-to-noise ratio (PSNR)

(11) *Discussion of Results*. Through the test result graph and evaluation index data, consider the pros and cons of the model in this paper. The thinking is mainly from the reasons for the best results, the reasons for the failure of the model, and the ability to further improve the network performance

#### 4. Experimental Verification and Result Analysis

4.1. *Evaluation Index*. The evaluation indicators used in this paper to evaluate the experimental effects of the model are PSNR and SSIM, respectively. Since the PSNR formula contains a mean square error (MSE) factor, first, introduce the MSE function. In Eq. (4), the symbol  $I$  represents a clean image, and  $K$  represents a noise image with a size of  $m \times n$ . In this experiment, the clean image  $I$  and the noise image  $K$ , respectively, represent the real image and the image generated by the model. MSE calculation formula is as follows

$$\text{MSE} = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i, j) - K(i, j)]^2. \quad (4)$$

The definition of PSNR is very similar to MSE. The unit of PSNR is db. In Eq. (5),  $(\text{MAX}_I)^2$  represents the maximum pixel value of the picture. PSNR calculation formula is as follows:

$$\text{PSNR} = 10 * \log_{10} \left( \frac{\text{MAX}_I^2}{\text{MSE}} \right). \quad (5)$$

SSIM is an important index used to evaluate images. It is composed of three measurement factors, namely, brightness

( $L$ ), contrast ( $C$ ), and structure ( $S$ ). The formulas for the three measurement factors are as follows:

$$L(x, y) = \frac{2u_x u_y + c_1}{u_x^2 + u_y^2 + c_1}, \quad (6)$$

$$C(x, y) = \frac{2\sigma_x \sigma_y + c_2}{\sigma_x^2 + \sigma_y^2 + c_2}, \quad (7)$$

$$S(x, y) = \frac{\sigma_{xy} + c_3}{\sigma_x \sigma_y + c_3}, \quad (8)$$

where  $u_x$  is the mean value of  $x$ ,  $u_y$  is the mean value of  $y$ ,  $\sigma_x^2$  is the variance of  $x$ ,  $\sigma_y^2$  is the variance of  $y$ , and  $\sigma_{x,y}$  is the covariance of  $x$  and  $y$ . The default value of  $K_1$  is 0.01, and the default value of  $K_2$  is 0.03. SSIM uses a special coupling method to couple the three measurement factors. The SSIM calculation formula is defined as follows:

$$\text{SSIM}(x, y) = [L(x, y)^\alpha \cdot C(x, y)^\beta \cdot S(x, y)^\gamma]. \quad (9)$$

For the convenience of calculation, set  $\alpha, \beta, \gamma$  to 1, then the original formula can be simplified. After Eqs. (6)–(8) is substituted into Eq. (9), the simplified SSIM formula is as follows:

$$\text{SSIM}(x, y) = \frac{(2u_x u_y + c_1)(2\sigma_{xy} + c_2)}{(u_x^2 + u_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}. \quad (10)$$

4.2. *Experimental Environment*. The experimental environment is the basic condition for conducting experiments. This section describes the experimental environment. The details are shown in Table 2.

4.3. *Experimental Results and Analysis*. To study the effectiveness and superiority of the Dual3D&PatchGAN model, four mainstream models were used as comparison algorithms in the experiment. Comparison algorithms mainly include medical image generation algorithm based on FCM [34], medical image generation algorithm based on WGAN [35], medical image generation algorithm based on CycleGAN transfer learning [36], and medical image generation algorithm based on Dual2D&PatchGAN model [37]. The

TABLE 3: PSNR indicators of each model.

Sample	FCM	CycleGAN	WGAN	Dual2DGAN	Dual3D&PatchGAN
1	12.4710	24.4511	19.9934	29.3316	28.7841
2	9.7536	27.6829	19.5638	25.4528	29.5963
3	13.5412	27.3729	21.5966	26.3986	31.7143
4	14.6819	25.4410	22.4536	23.5241	32.4196
5	12.2365	22.3596	27.5682	27.3351	27.5655
6	14.6512	28.5151	22.5510	29.6637	29.5122
7	17.7714	28.5974	22.5809	23.4510	33.4243
8	14.2287	24.2586	23.4814	27.5968	29.5399
9	15.6632	28.6624	21.1940	28.5129	30.0019
Means	13.8887	26.3712	22.3314	26.8074	30.2842
Variance	5.1887	5.3292	5.4789	5.2985	3.4762

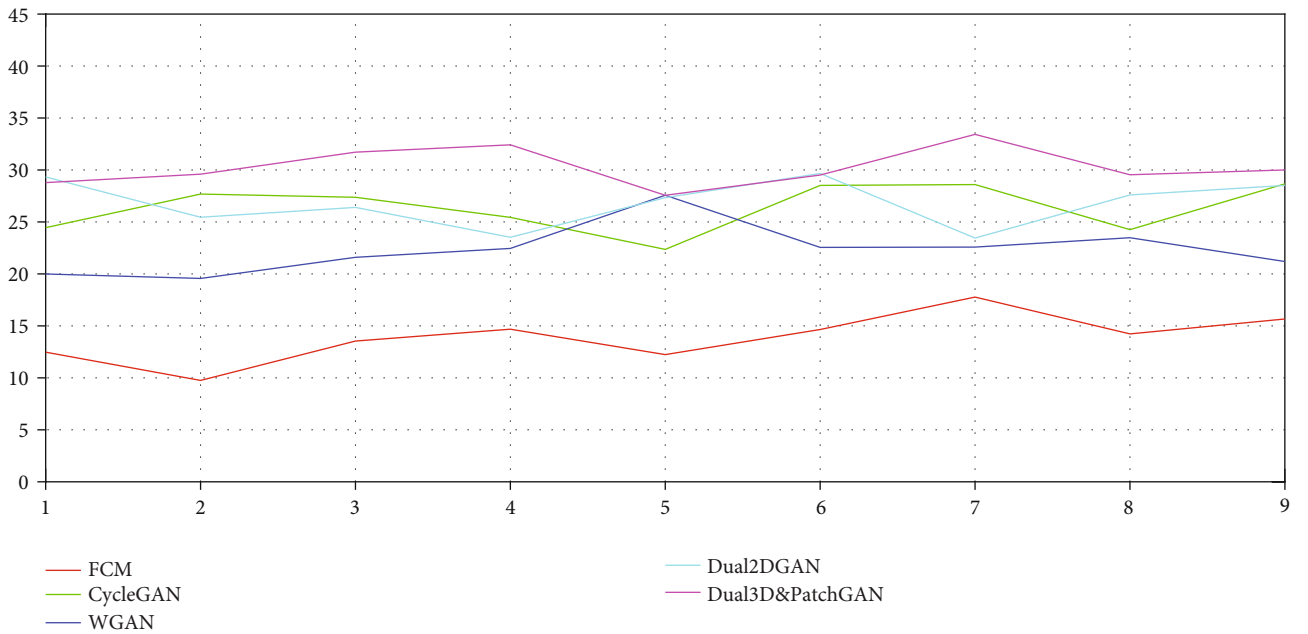


FIGURE 5: Comparison of PSNR indicators.

experimental data is the brain image data downloaded on the brainweb website. A total of 9 samples of brain CT images and magnetic resonance (MR) images were downloaded as experimental data sets. In order to analyze the effectiveness of the used model more objectively, the evaluation index based on the results generated by the five models is calculated. Analyze the image generation effect of each model through evaluation indicators.

Before each model is trained, the data must be cleaned first to eliminate images with artifacts, blur, incompleteness, and other undesirable factors and select images with better quality. The number of selected images should be as many as possible. Place the selected images according to the CT domain and MR domain, respectively. Due to the limited sample data in this experiment, data enhancement is performed on the selected CT domain and MR domain images. Data enhancement is to allow limited data to gen-

erate more data value. In this experiment, strategies such as flipping, rotation, cropping, deformation, and zooming were used to enhance the original data to ensure the learnability of the experimental data as much as possible. After the data required for the model is processed, the model can be trained.

The two data domains in this experiment are CT domain and MR domain images of 9-bit samples. Since the Dual3D&PatchGAN model is based on the idea of transfer learning, the data sets do not need to be paired, so when selecting the data set, only high-quality images in two domains need to be selected. The model needs to be trained multiple times. The data details during training can be viewed through Tensorboard and the output console, and the model parameters can be modified in time to improve the generalization ability of the model. After many sessions of training, a stable and well-performing network model is



TABLE 4: SSIM indicators of each model.

Sample	FCM	CycleGAN	WGAN	Dual2DGAN	Dual3D&PatchGAN
1	0.5398	0.6101	0.6705	0.6305	0.8252
2	0.4975	0.7932	0.7732	0.7021	0.7895
3	0.5534	0.7205	0.6250	0.6552	0.8102
4	0.5801	0.7022	0.7532	0.7306	0.8521
5	0.5003	0.6352	0.6609	0.8051	0.8012
6	0.5908	0.6010	0.6920	0.7334	0.7965
7	0.4431	0.6905	0.6679	0.6988	0.7849
8	0.4806	0.5405	0.7001	0.7005	0.8502
9	0.5004	0.6609	0.7203	0.6967	0.8106
Means	0.5207	0.6616	0.6959	0.7059	0.8139
Variance	0.0024	0.0056	0.0022	0.0025	0.0006

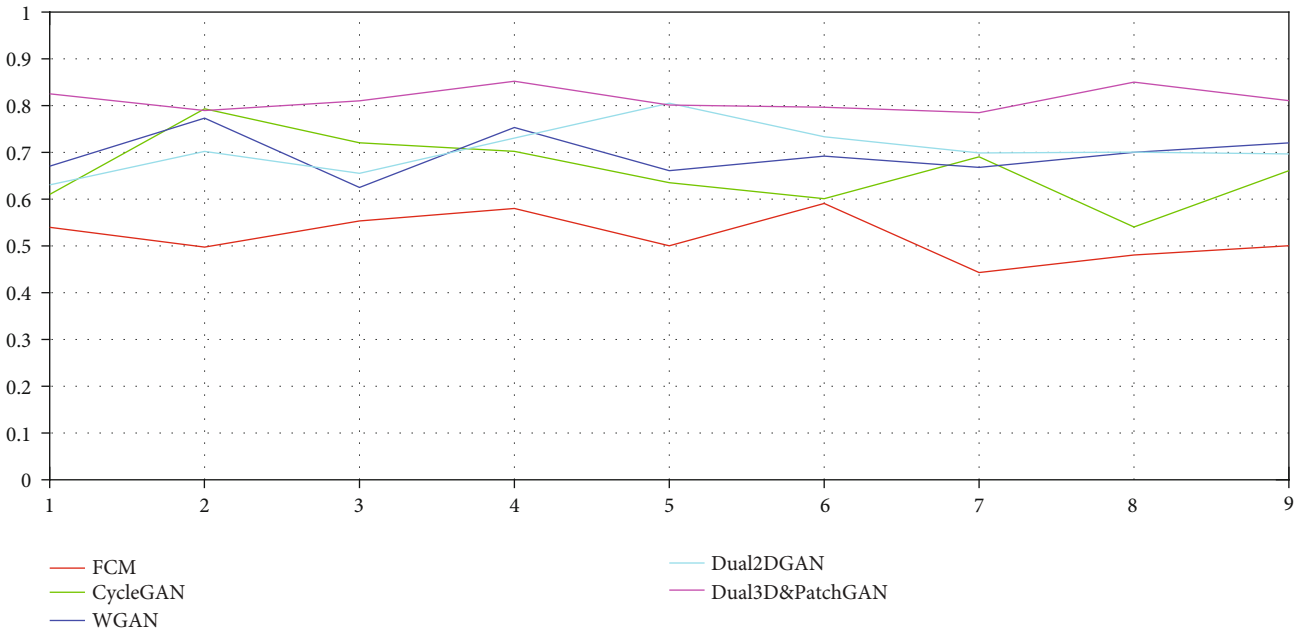


FIGURE 6: Comparison of SSIM indicators.

gradually established for subsequent operations. During each training, the CT and MR images of 8 samples are used as the training set, and the CT and MR images of the remaining sample are used as the test set for simulation image generation. Each test needs to store the test results and make an annotation. By checking the quality of the test image, the network model parameters can be dynamically adjusted, and the best parameter model can be found step by step. In practical applications, the sample usually has a medical image, such as an MR image. However, the lesion requires one or more other medical images as auxiliary means. In order to shorten the detection time and make the body receive as little radiation as possible, the use of a computer for image analog conversion is a feasible medical aid. Table 3 and Figure 5 show the PSNR of the images generated by each model. Table 4 and Figure 6 show the SSIM of the images generated by each model.

From the results of the two evaluation indicators, it can be analyzed that the FCM model has the worst effect and is not competent for the task of image generation. Among several network models, the generation performance of the network model used in this article is significantly higher than other network models. The stability of Dual3D&PatchGAN model is significantly better than other models. This is because the Dual3D&PatchGAN model is aimed at a 3D model, and the result generated each time is a three-dimensional block, which is often more accurate for the model result. Moreover, the transfer idea in the model effectively utilizes the information of the source domain data, thereby improving the generation performance. The Dual3D&PatchGAN model is based on the improvement of Dual2DGAN. The Dual3D&PatchGAN model is no longer limited to a 2-dimensional model, that is, it is no longer limited to a flat image on the image but can also be a 3-

dimensional image. Because this model adds the PatchGAN module, when calculating Softmax, the whole image is no longer calculated, but the image is divided into several small blocks. Each small block is calculated separately, summarized, and averaged to obtain the Softmax result, thus ensuring the accuracy of the calculation result. In addition, the model also adds a training sequence disturbance module and a random cropping module, which also disrupts the training sequence of the model to ensure the randomness of training. In the initial network model training, the learning rate is chosen to be a smaller fixed value, which is set to  $2E-4$  in this article. By looking at the Tensorboard in the Tensorflow framework, it is found that the loss function of the model is difficult to reduce, and it is often oscillating in the final stage. This means that it is difficult to obtain a stable training model. So this article sets the learning rate to  $1E-5$ , and as the training time increases, the loss function gradually decreases. It can be seen from the experimental results that the images generated by DualGAN3D&PatchGAN have the best results.

## 5. Conclusion

In this paper, it is mainly used to generate medical images based on the idea of transfer learning. At the same time, several methods and ideas to solve this problem are discussed. Compared with the unique idea of fuzzy C-means and the traditional GAN model, Dual3D&PatchGAN based on 3D convolution is often more valuable, and it reduces the requirements for data sets. On the other hand, this method does not require the one-to-one pairing of two medical images. Due to the excellent performance of the GAN model, many scholars often use the GAN model to generate high-quality medical images. Coincidentally, this unique idea is also used in this paper, but the idea of transfer learning, 3D convolution, and PatchGAN is added to the network. We should further improve network performance, regarding the generation of medical images in multiple fields, while ensuring the quality of medical images, how to convert low-information images into high-information medical images is a question worth discussing, and we hope that some more exciting works will appear in the future.

## Data Availability

The dataset analyzed for this study can be found in this link [<https://brainweb.bic.mni.mcgill.ca/>].

## Conflicts of Interest

The authors declare no conflicts of interest.

## Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 61772241, in part by the 2018 Six Talent Peaks Project of Jiangsu Province under Grant XYDXX-127.

## References

- [1] S. Dehghan-Dehnavi, M. Fotuhi-Firuzabad, M. Moeini-Aghtaie, P. Dehghanian, and F. Wang, "Decision-making tree analysis for industrial load classification in demand response programs," *IEEE Transactions on Industry Applications*, vol. 57, no. 1, pp. 26–35, 2020.
- [2] M. Shaheen, T. Zafar, and A. K. Sajid, "Decision tree classification: ranking journals using IGIDI," *Journal of Information Science*, vol. 46, no. 3, pp. 325–339, 2020.
- [3] M. Sun and R. Yang, "An efficient secure k nearest neighbor classification protocol with high-dimensional features," *International Journal of Intelligent Systems*, vol. 35, no. 11, pp. 1791–1813, 2020.
- [4] J. Shi and L. Yang, "A climate classification of China through k-nearest-neighbor and sparse subspace representation," *Journal of Climate*, vol. 33, no. 1, pp. 243–262, 2020.
- [5] Z. Ebrahimpour, W. Wan, A. S. Khojine, and L. Hou, "Twin hyper-ellipsoidal support vector machine for binary classification," *IEEE Access*, vol. 8, pp. 87341–87353, 2020.
- [6] Z. L. Yang, S. Y. Zhang, Y. T. Hu, Z. W. Hu, and Y. F. Huang, "VAE-Stega: linguistic steganography based on variational auto-encoder," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 880–895, 2021.
- [7] M. Sadeghi, S. Leglaive, X. Alameda-Pineda, L. Girin, and R. Horaud, "Audio-visual speech enhancement using conditional variational auto-encoders," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 1788–1800, 2020.
- [8] I. Goodfellow, J. Pouget-Abadie, and M. Mirza, *Generative Adversarial Nets. Advances in Neural Information Processing Systems*, ACM, New York, 2014.
- [9] B. W. Tseng and P. Y. Wu, "Compressive privacy generative adversarial network," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 2499–2513, 2020.
- [10] B. Zhu, J. Jiao, and D. Tse, "Deconstructing generative adversarial networks," *IEEE Transactions on Information Theory*, vol. 66, no. 11, pp. 7155–7179, 2020.
- [11] H. Phan, I. V. McLoughlin, L. Pham et al., "Improving GANs for speech enhancement," *IEEE Signal Processing Letters*, vol. 27, pp. 1700–1704, 2020.
- [12] W. Peng, Y.-H. Dai, H. Zhang, and L. Cheng, "Training GANs with centripetal acceleration," *Optimization Methods and Software*, vol. 35, no. 5, pp. 955–973, 2020.
- [13] N. Dong, R. Trullo, J. Lian, C. Petitjean, and D. Shen, "Medical image synthesis with context-aware generative adversarial networks," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 417–425, Springer, Cham, 2017.
- [14] G. Yang, S. Yu, H. Dong, and G. Slabaugh, "DAGAN: deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1310–1321, 2018.
- [15] L. Hou, A. Agarwal, D. Samaras, T. M. Kurc, R. R. Gupta, and J. H. Saltz, *Unsupervised Histopathology Image Synthesis*, Advances in Neural Information Processing Systems, 2017.
- [16] P. Adarsh, B. Arunav, P. Uddipta, A. Sahil, and B. Manali, "Implementation of DCGAN to Generate Gamocha Design Patterns," *Lecture Notes in Electrical Engineering*, vol. 686, pp. 831–838, 2020.

- [17] E. Denton, S. Chintala, A. Szlam, and R. Fergus, *Deep Generative Image Models Using a Laplacian Pyramid of Adversarial Networks*, Advances in neural information processing systems, 2015.
- [18] L. Zhang, J. Zhao, X. Ye, and Y. Chen, "Cooperation: a new force for boosting generative adversarial nets with dual-network structure," *IET Image Processing*, vol. 14, no. 6, pp. 1073–1080, 2020.
- [19] M. Rubin, O. Stein, N. A. Turko et al., "TOP-GAN: stain-free cancer cell classification using deep learning with a small training set," *Medical Image Analysis*, vol. 57, pp. 176–185, 2019.
- [20] L. Teng, Z. Fu, and Y. Yao, "Interactive translation in echocardiography training system with enhanced cycle-GAN," *IEEE access*, vol. 8, pp. 106147–106156, 2020.
- [21] S. U. Dar, M. Yurt, L. Karacan, A. Erdem, E. Erdem, and T. Cukur, "Image synthesis in multi-contrast MRI with conditional generative adversarial networks," *IEEE Transactions on Medical Imaging*, vol. 38, no. 10, pp. 2375–2388, 2019.
- [22] F. Amirjavid, S. Barak, and H. Nemati, "A fuzzy paradigmatic clustering algorithm," *IFAC-PapersOnLine*, vol. 52, no. 13, pp. 2360–2365, 2019.
- [23] Q. T. Bui, B. Vo, V. Snasel et al., "SFCM: a fuzzy clustering algorithm of extracting the shape information of data," *IEEE Transactions on Fuzzy Systems*, vol. 29, no. 1, pp. 75–89, 2021.
- [24] M. Selvi, K. S. Santhosh, S. Ganapathy, A. Ayyanar, K. Nehemiah, and K. Arputharaj, "An energy efficient clustered gravitational and fuzzy based routing algorithm in WSNs," *Wireless Personal Communications*, vol. 116, no. 1, pp. 61–90, 2021.
- [25] O. Kasmi, A. Baina, and M. Bellafkih, "Fuzzy logic based clustering algorithm for management in critical infrastructure," *Cluster Computing*, vol. 24, no. 1, pp. 433–458, 2021.
- [26] L. C. N. Kamila and I. Ranggadara, "Fuzzy sugeno algorithm for clustering document management," *International Journal Of Advanced Trends In Computer Science and Engineering*, vol. 9, no. 1, pp. 26–30, 2020.
- [27] Z. Peng, W. Zhang, N. Han, X. Fang, P. Kang, and L. Teng, "Active transfer learning," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 4, pp. 1022–1036, 2019.
- [28] K. Weimann and T. O. F. Conrad, "Transfer learning for ECG classification," *Scientific Reports*, vol. 11, no. 1, pp. 1–12, 2021.
- [29] S. Wang and N. Xi, "Calibration of haptic sensors using transfer learning," *IEEE Sensors Journal*, vol. 21, no. 2, pp. 2003–2012, 2021.
- [30] C. X. Ren, D. Q. Dai, K. K. Huang, and Z. R. Lai, "Transfer learning of structured representation for face recognition," *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 5440–5454, 2014.
- [31] T. Russ, S. Goertler, and A. K. Schnurr, "Synthesis of ct images from digital body phantoms using cyclegan," *International Journal of Computer Assisted Radiology and Surgery*, vol. 14, no. 10, pp. 1741–1750, 2019.
- [32] K. B. Kancharagunta and D. Shiv Ram, "PCSGAN: perceptual cyclic-synthesized generative adversarial networks for thermal and NIR to visible image transformation," *Neurocomputing*, vol. 413, pp. 41–50, 2020.
- [33] X. Qu, X. Wang, Z. Wang, L. Wang, and L. Zhang, "Perceptual-DualGAN: perceptual losses for image to image translation with generative adversarial nets," in *Proceedings of The International Joint Conference on Neural Networks*, Rio de Janeiro, Brazil, 2018.
- [34] N. R. Pal and J. C. Bezdek, "On cluster validity for the fuzzy c-means model," *IEEE Transactions on Fuzzy Systems*, vol. 3, no. 3, pp. 370–379, 1995.
- [35] Z. Yang, Y. Chen, Z. Le, F. Fan, and E. Pan, "Multi-source medical image fusion based on wasserstein generative adversarial networks," *IEEE Access*, vol. 7, pp. 175947–175958, 2019.
- [36] V. Sandfort, K. Yan, P. J. Pickhardt, and R. M. Summers, "Data augmentation using generative adversarial networks (cycle-gan) to improve generalizability in ct segmentation tasks," *Scientific Reports*, vol. 9, no. 1, p. 16884, 2019.
- [37] Z. Yi, H. Zhang, and P. T. M. Gong, "DualGAN: unsupervised dual learning for image-to-image translation," *IEEE International Conference on Computer Vision*, vol. 2017, pp. 2868–2876, 2017.

## Research Article

# Network Intrusion Detection Based on an Improved Long-Short-Term Memory Model in Combination with Multiple Spatiotemporal Structures

Xiaolong Huang 

*School of Information Engineering, Baise University, Baise 533000, China*

Correspondence should be addressed to Xiaolong Huang; [hsl@bsuc.edu.cn](mailto:hsl@bsuc.edu.cn)

Received 8 December 2020; Revised 29 December 2020; Accepted 13 April 2021; Published 24 April 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Xiaolong Huang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Aimed at the existing problems in network intrusion detection, this paper proposes an improved LSTM combined with spatiotemporal structure for intrusion detection. The unsupervised spatiotemporal encoder is used to intelligently extract the spatial characteristics of network traffic data samples. It can not only retain the overall/nonlocal characteristics of the data samples but also extract the most essential deep features of the data samples. Finally, the extracted features are used as input of the LSTM model to realize classification and identification for intrusion samples. Experimental verification shows that the accuracy and false alarm rate of the intrusion detection model based on the neural network are significantly better than those of other traditional models.

## 1. Introduction

With the continuous development of the Internet, the network has been integrated into all aspects of people's daily life, so network security has become a problem that network users have to face [1]. The importance of network security has risen to the height of a national strategy. If the network data is attacked, it will have an irreversible impact on the efficiency and security of the enterprise [2]. IDS is defined as an intrusion detection system [3] that can extract and analyze the characteristics of the input data and intercept the detected abnormal data, which greatly improves the security of the system. IDS refers to the establishment of a rule base based on existing knowledge or the training of abnormal behavior characteristics to detect malicious attacks such as computer worms and Trojan, so as to maintain information security. An accurate and stable intrusion detection system is very important to a network security system [4]. Therefore, an intrusion detection system has important research significance.

Intrusion detection is an indispensable defense line in the security system. It collects information from several key nodes in the computer network system, checks whether there are any violations of security policies and signs of attack in the network, identifies threats in the network, and generates alarms, so as to provide real-time protection for internal attacks, external attacks, and misoperations [5]. In fact, intrusion detection is to classify the network traffic packets into two or more categories, dividing each network connection into normal behavior or abnormal attack or further distinguishing which kind of attack it belongs to through a multi-object classification model.

However, with the continuous development of science and technology and the gradual opening of the public network, the operation efficiency and real-time situational awareness efficiency of many facilities and equipment are also improved simultaneously. The scale of the Internet is becoming larger and larger, and the structure is more and more complex [6]. Due to the long-term experience of the attacker, the attack mode is different from the conventional

attack mode in the past, which is more intelligent and complex. At present, many scholars have studied intrusion detection models.

In recent years, many machine learning methods have been widely used to identify various types of attacks in the network and the other applications [7–12]. However, most traditional machine learning algorithms belong to shallow learning, and shallow learning algorithms rely on the construction of model and selection of data features [7]. They cannot efficiently classify large-scale data in the real network environment. Shallow learning cannot meet the requirements of intelligent analysis of massive data and prediction of high-dimensional learning [8]. Intrusion detection based on shallow models, such as traditional data mining and machine learning methods, is difficult to effectively detect various new types of attacks. Deep learning has the potential to extract better representations from massive data to design better detection models [9].

Deep learning methods can improve the overall performance of intrusion detection systems, which is a popular research direction for many researchers, and relevant research results are also emerging in endlessly [13]. He et al. proposed an improved deep learning method for flow-based anomaly detection, and experiment results have showed that deep learning can be applied to software-defined network anomaly detection [14]. Umer et al. [15] proposed a deep belief network-based intrusion detection model to classify network connections and verified the effectiveness of the proposed method on the NSL-KDD dataset. However, these literatures mainly focus on using deep learning methods as part of a pretraining model and then using traditional methods such as decision trees and SVM for classification. Afterwards, some researchers have successively use deep learning methods to build intrusion detection models. Tan et al. [16] proposed an intrusion detection model constructed by a three-layer RNN and proved through experiments that the model can improve the performance of intrusion detection. However, the various layers in the network are partially connected. That is to say, the simplified RNN model does not reflect the ability to learn the deep feature of the data, and the classification performance of the model has not been analyzed in the binary classification. Subsequently, Liang et al. [17] proposed an intrusion detection model constructed by a fully connected RNN, where the model was studied for binary-class and five-class classification on the NSL-KDD dataset. The experimental results showed that the intrusion detection model based on RNN is better than traditional machine learning, which can get a higher detection rate. However, this model does not remove redundant features when training the RNN classification network and is doped with a lot of data noise, so the final classification performance of the model is not ideal. Deep learning methods have natural advantages over shallow learning methods in the face of large-scale data, where better detection results can usually be obtained in intrusion detection. In addition, as the dimensionality of data features increases, the hidden layer structure of deep neural networks will become more complex, and the difficulty of model training will continue to increase. Moreover, the features

extracted from the network connection data are often redundant, which will also make the detection rate of the model reduced [18].

Literature [19] proposed an intrusion detection model based on a deep autoencoding network. Although it improves the classification accuracy and detection rate, it ignores the sequence characteristics of intrusion data. Literature [20] proposed the combination of a asymmetric convolutional autoencoder and supports a vector machine to be applied to intrusion detection systems, which significantly reduced the training time, but the detection time was longer and the robustness was not good; literature [21] proposed the combination of the encoder which extracts data features and the extreme learning machine which classifies quickly and effectively, but the disadvantage is that it is easy to fall into overfitting; the PCA-LSTM algorithm proposed in literature [22] can effectively remove the noise information in the sample data. However, when the amount of data is large, the accuracy of data extraction is low; literature [23] proposed a novel hybrid algorithm based on the PCA-ANN model, and the Artificial Neural Network (ANN) has great advantages in reducing training time, unsupervised learning, and highly nonlinear approximation capabilities, but it is easy to fall into a local minimum. In addition, the feed-forward neural network in ANN does not have the function of remembering and using long-term information dependence. In summary, in view of the current network attacks that are intelligent, complicated, and concealed, the amount of data is large, and complexity and feature dimensions are high; this paper proposes a deep learning-based LSTM neural network intrusion detection model. It can solve the problem of gradient disappearance and gradient explosion of the traditional recurrent neural network (RNN). Traditional shallow learning methods cannot detect current intelligent attack methods, such as hiding in the object-host for 1 to 2 years. LSTM has the advantage of using feature data related to long-term dependence and long-term span and has stronger training and detection capabilities than recurrent neural networks. With the rapid development of Internet technology in the current era of big data, there are not only more attacks but also greatly increased network traffic compared with the traditional Internet. Therefore, intrusion data presents the characteristics of large samples and high dimensions. When the PCA feature extraction model is applied to a large number of data samples, the problem of incomplete feature expression will lead to an increase in the false-positive rate of detection [24]. Deep learning has more advantages in processing large samples and high-dimensional data. SDAE adds noise to data information on the basis of the autoencoder model, which enhances the robustness of the input layer of the autoencoding network. Compared with DAE, it adds lost packet technology. Robustness of feature cascade between autoencoding networks has also improved the robustness of intrusion detection. Therefore, some scholars adopted the stacked denoise autoencoder to perform the spatial dimensionality reduction reconstruction for high-dimensional data.

Aimed at the existing problems in network intrusion detection, this paper proposes an improved LSTM combined with spatiotemporal structure for intrusion detection. The

unsupervised spatiotemporal encoder is used to intelligently extract the spatial characteristics of network traffic data samples. It can not only retain the overall characteristics of the data samples but also extract the deep features of the data samples. Deep learning extracts the spatial feature of network traffic data and extracts more complex features iteratively on the basis of preserving the overall characteristics of the data. Finally, the extracted features are used as input of the LSTM model to realize classification and identification for intrusion samples. The model is trained and tested by using the NSL-KDD dataset which is more in line with the data characteristics of the new era than the KDDCUP dataset. Experimental verification shows that the accuracy and false alarm rate of the intrusion detection model based on the neural network are significantly better than those of other traditional models.

## 2. Long-Short-Term Memory Network

The core of deep learning is to train the weight parameter matrix and bias parameter deeply through multiple neural networks, so as to minimize the error function between the calculated value and the real value in the deep learning model. The derivative of the error function with respect to the weight and bias parameters is obtained, and the weight and bias parameters under the minimum value are determined by using the change trend of the error function. The cost of calculating the loss function of each parameter is very high. The error backpropagation has been regarded as a gradient descent method, which can usually be used to solve the gradient problem. The global gradient solution is transformed into the local gradient solution, which simplifies the calculation process [24].

The current output of a sequence of recurrent neural networks (RNN) is affected by both the current input and the previous output. It has the function of “memory” to the information in the front of the network. When recursing in time, it can be regarded as a limited multilayer deep learning network [25]. For RNN, the basic function of each hidden state layer is to memorize data and add new information to each layer through each iteration so that the information is passed down.

Although RNN can effectively deal with nonlinear time series data, there are still two problems: (1) due to gradient vanishing and gradient explosion, RNN cannot process long time series data and (2) the training of the RNN model needs to determine the intercept length, and its optimal parameters are difficult to obtain by experience. LSTM can effectively solve these problems. LSTM and RNN have the same input and output, but the difference is the internal structure of the hidden layer. LSTM is an improved recurrent neural network and is mainly to overcome the defect of RNN that is difficult to deal with long-distance dependence in practical application, which is the most popular RNN at present. Their structural differences are shown in Figure 1. LSTM has made milestone achievements in many fields such as speech recognition, image description, and natural language processing.

In addition to the external RNN recurrent, the LSTM recurrent network also introduces an internal self-recurrent. The weight in the self-recurrent depends on the context, so

that the neural network can selectively forget the old state and ensure the continuous flow of the gradient. Therefore, LSTM is not a simple nonlinear model with element by element operation after transforming input and recurrent units but a complex nonlinear system of a gating unit system including more parameters and control information flow. Its gating unit system consists of three parts: forgetting gate, input gate, and output gate. Its structure is shown in Figure 2.

The weight of the internal self-recurrent is mainly controlled by the forgetting gate. Its activation function adopts the tanh function, so that the value of the weight parameter ranges from 0 to 1, and its output shape is the same with  $C_{t-1}$ . They are multiplied point by point to determine the old state of forgetting; the input gate is also called the memory gate, which is used to determine the new state of memory. The output gate is the result of combining the forgetting gate and the memory gate to calculate the value of the next hidden state. The state updating method of the internal recurrent in LSTM is shown and written as follows.

$$f_t = \sigma \left( b_f + \sum U_f x_t + \sum W_f h_{t-1} \right), \quad (1)$$

$$g_t = \sigma \left( b_g + \sum U_g x_t + \sum W_g h_{t-1} \right), \quad (2)$$

$$q_t = \sigma \left( b_o + \sum U_o x_t + \sum W_o h_{t-1} \right), \quad (3)$$

$$C_t = f_t C_{t-1} + \tanh \left( b + \sum U_i x_t + \sum W_i h_{t-1} \right) g_t, \quad (4)$$

$$h_t = \tanh (C_t) q_t, \quad (5)$$

where  $x_t$  is the current input,  $h_t$  and  $h_{t-1}$  are the hidden state of the  $t$ -th time and the  $(t-1)$ -th time,  $f_t$  is the forgetting gate,  $g_t$  is the input gate,  $q_t$  is the output gate,  $U$ ,  $W$ , and  $B$  are the weight parameters in the self-recurrent,  $\sigma(\cdot)$  and  $\tanh(\cdot)$  are sigmoid and hyperbolic tangent activation functions, respectively. The LSTM model training process adopts the time backpropagation algorithm, which is roughly divided into the following four steps [26].

- (1) The forward propagation algorithm in the LSTM model is obtained by using the internal self-recurrent state update method in the LSTM model, and the output value of the single module is calculated
- (2) From two directions of the time and network layer, the error term of each LSTM is calculated reversely
- (3) The gradient of each weight is calculated for the error term
- (4) The gradient-based optimization algorithm is applied to update the weight.

The commonly used parameter updating optimization methods include stochastic gradient descent, momentum, AdaGrad, RMSProp, and Adam [27]. In this paper, Adam, namely, adaptive momentum estimation algorithm, is selected as the optimization algorithm. The algorithm combines the advantages of the SGD algorithm and momentum

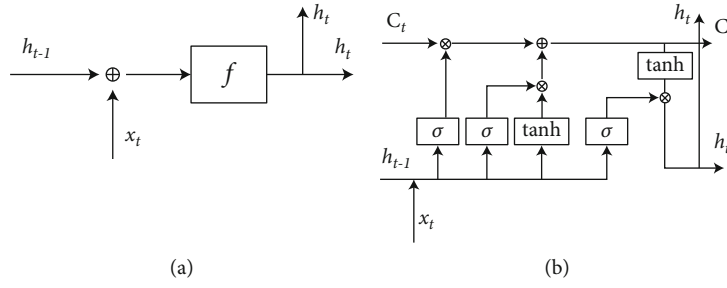


FIGURE 1: Hidden layer internal structure diagram: (a) RNN; (b) LSTM.

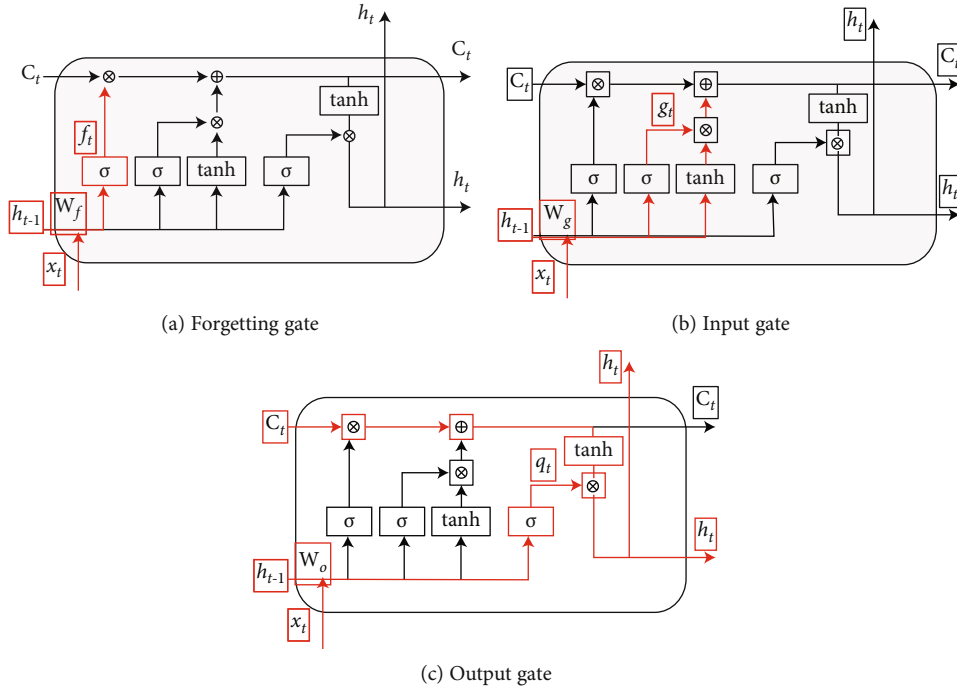


FIGURE 2: Schematic diagram of LSTM gate structure.

algorithm to search for the parameter space efficiently, and it can be used to perform the “bias correction” of the super parameters, which is less dependent on the initial value and takes less computer resources in the calculation process.

### 3. Multiple Spatiotemporal Models for LSTM

In order to make the characteristics of each node in the network structure represent different variables, the spatiotemporal characteristics extracted by each node can be used to independently predict different variables [28]. Therefore, after the network is trained, the characteristics of each node can correspond to different variables. This learning method can construct multiple independent channels, and each channel represents a variable and is used to learn each variable. According to the above ideas, it is possible to realize the mining of the spatiotemporal characteristics in the Internet. When the characteristics of each node not only correspond to different variables but also contain spatiotemporal information, the convolution operation is performed on the

basis of the spatiotemporal relationship in the Internet, and then, the extracted features can have both the temporal and spatial characteristics of the process. In the process of training, the network can coordinate the learning of temporal and spatial relationships to obtain the intrusion characteristics of spatiotemporal information in the fusion process [29]. Therefore, the soft measurement application can be realized on the basis of collaborative learning of process spatiotemporal characteristics. According to the above ideas, this paper adopts a multichannel network structure to realize spatiotemporal collaborative learning for the intrusion detection system.

In the training process of the deep network, the LSTM and the spatiotemporal model can cooperate with each other in the learning of time series characteristics and spatial characteristics [29] and have the advantages of both. First, the process variables are input into  $p$  separate channels, and each channel represents a variable, and the LSTM layer is used to extract the timing features for different variables; then, the timing features extracted by each channel are used as nodes

in the spatiotemporal structure. It is worth noting that although the feature of each node does not directly represent the variable but is extracted from the process variable, it can be targeted to the prediction of each variable after the reverse training of the network. Combined with the adjacency matrix learned from sparse coding, the use of cross-channel spatiotemporal convolution operations can fuse the timing characteristics of different variables in different channels of spatiotemporal structure and associate channels with strong spatial relationships. Therefore, the extracted spatiotemporal features can reflect the inherent characteristics of the process. Since the operation of the spatiotemporal convolutional layer merges the timing characteristics in the related channels, the fused spatiotemporal characteristics can still retain the timing characteristics of the original characteristics; finally, after each channel is processed by two FC layers, the prediction results of each variable are obtained. Through the design of the multichannel network structure, the characteristics of the LSTM and spatiotemporal module can learn the spatiotemporal characteristics which reflect the process characteristics in the process of network training. Each variable can not only learn independently in its own channel but also be effectively correlated to make the learning process of different variables independent and shared.

Therefore, a network intrusion detection model based on multiple spatiotemporal models in LSTM is proposed in this paper [30]. In fact, the essence of the intrusion detection is a classifier model, in which the network flow mixed with abnormal data can be detected. In this paper, a spatiotemporal module is added to the LSTM structure of the classifier to extract features from the data. The process can retain the overall features of the data, and the more complex features are iteratively extracted from the low-level features, thus reducing the feature dimension of the original data. After the spatiotemporal model completes pretraining, the state vectors of each forgetting gate are combined to construct a multilayer neural network. In order to detect abnormal intrusion data, the LSTM with discriminative ability is used as the output layer for network abnormal intrusion detection, identification features, and classification, so as to output the classification detection results of traffic flow data. As shown in Figure 3, the internal loop structure of the LSTM neural network is expanded in time. The input of each time step is the output feature vector sequence of the spatiotemporal model, where  $x_t$ ,  $x_{t-1}$ , and  $x_{t+1}$  represent the current state, the previous state, and the next state, respectively. LSTM gradually transfers the memory state backwards in time order by controlling its internal gate structure. The interface between spatiotemporal and LSTM neural networks requires that the input of the LSTM neural network is a sequence of feature vectors. In other words, it is composed of a collection of feature vectors at consecutive  $T$  time steps [29]. Therefore, the input sequence of LSTM must be constructed before training. The construction method is as follows: let  $x_t$  be the feature vector of the  $t$ -th time step; then, the first input sequence is  $\{x_1, x_2, \dots, x_T\}$ , and the second is  $\{x_2, x_3, \dots, x_{T+1}\}$ . By analogy, all interface input sequences of the spatiotemporal model and LSTM are obtained.

In order to better adapt the anomaly intrusion detection, we built a unidirectional 3-layer LSTM stacked neural network. As shown in Figure 3, the model is a stack of three LSTM layers.  $\{x_0, x_1, \dots, x_N\}$  is the input data after preprocessing. The final output layer is a classification network using the softmax function. The softmax function is essentially a form of probability distribution of neuron output, and the number of nodes in the last output layer is equal to the number of the classification task. The final output data uses one-dimensional arrays  $[0, 1]$  and  $[-1, 0]$  to represent intrusion traffic and normal traffic. In actual prediction, we use the subscript of the maximum value in the one-dimensional array to represent the prediction result; for example,  $[0.0134871, 0.9875801]$  represents intrusion traffic and  $[0.9965685, 0.0031123]$  represents normal traffic. The increase in network capacity can easily lead to overfitting of the model. In order to prevent overfitting and improve the generalization performance of the model, we use recurrent dropout regularization to reduce overfitting in the training process. In other words, the input unit of a certain layer is randomly set to 0 with a certain probability, and the purpose is to break the accidental correlation in the training data of this layer [30]. On the input data of the model, the dataset is transformed into the form of input data required by the model through feature extraction and preprocessing of the imported dataset. Each vector contains data with  $N$  feature values, where the batch size is 60 and the time step is 100 in the model.

The LSTM layer and spatiotemporal layer in the above multichannel network structure are defined by equations (1)–(5), and the FC layer is defined by

$$H^{(l)} = \sigma \left( H^{(l-1)} W^{(l)} + b^{(l)} \right), \quad (6)$$

where  $H^{(l)}$  is the implicit feature of layer  $l$ ,  $W^{(l)}$  and  $b^{(l)}$  are weight and bias parameters, respectively, and  $\sigma(\cdot)$  is the ReLu activation function. The loss function of our proposed model is defined by formula (7). The model parameters in each layer are optimized by gradient descent, where  $Y$  and  $\hat{Y}$  represent the real and predicted values of variables, respectively. In order to avoid overfitting in training, the  $l^2$  norm regularization term for model parameters is introduced, and the weight coefficient of the regularization term is  $\gamma$ .

$$\text{loss}(Y, \hat{Y}) = \sum_{i=1}^m \sum_{j=1}^p (Y_{ij} - \hat{Y}_{ij})^2 + \gamma \|W\|^2. \quad (7)$$

In practical application, first of all, we need to use the variable matrix in the training set and use the sparse coding to learn the spatiotemporal structure among variables and keep the spatiotemporal structure unchanged in the process of subsequent model training and testing. Our proposed model can be obtained by constructing the multichannel network structure. Through the multichannel network structure designed by our proposed model, LSTM and spatiotemporal modules are successively used to mine and learn the temporal and spatial characteristics of the intrusion detection process,



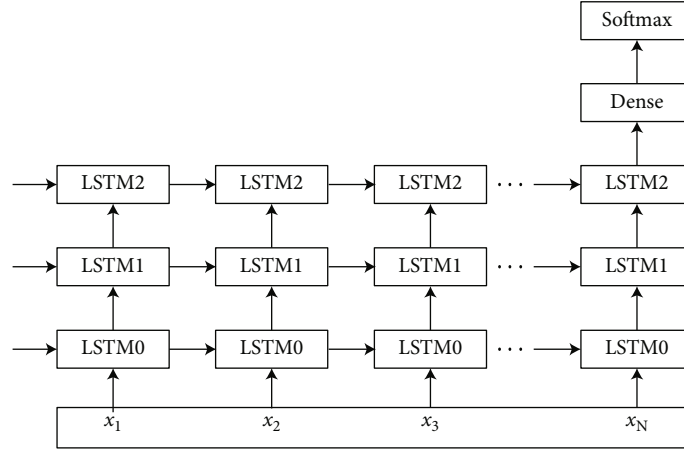


FIGURE 3: Improved LSTM combined with spatiotemporal structure.

and the features that can represent the internal temporal and spatial properties of the process are obtained for the intrusion detection system. Compared with the traditional model which only considers the temporal sequence characteristics of the process, our proposed model makes full use of the advantages of LSTM and spatiotemporal module and effectively realizes the intrusion detection on the basis of spatiotemporal collaboration.

#### 4. Experimental Results and Analysis

**4.1. Experimental Environment.** In order to verify the abnormal network attack detection model based on the multispatiotemporal and long-short-term memory model in this paper, a simulation experiment environment is built [31]. The experiment uses the Keras2.2.4 deep learning framework based on tensorflow-GPU1.13 for simulation, the operating system is Windows 10, the Intel i5-6300HQ 4-core processor is CPU, the memory size is 8 G, and the NVIDIA GTX960 graphics card is used to accelerate the running speed of the model.

**4.2. Evaluation Index.** In this experiment, four indicators are constructed to evaluate the advantages and disadvantages of the model, including the accuracy rate (ACC), false-positive rate (FPR), false-negative rate (FNR), and prediction rate (DR). Through the evaluation indexes, we can get the advantages and disadvantages of the model and then adjust the model parameters until the evaluation index is optimal [31], which means that the model is optimal at this time. As shown in Table 1,  $T^+$  represents that the predicted result is intrusion attack, namely, the number of successful predicted samples;  $T^-$  represents the number of samples predicted as normal and the actual value as the number of samples with successful normal prediction. Therefore, we can use the ACC to denote the accuracy:  $ACC = (T^- + T^+) / (T^- + T^+ + F^- + F^+)$ .  $F^-$  represents that the predicted result is normal, but the actual value is attacked; namely, the prediction fails. Therefore, the system fails to detect the number of samples which can be denoted as the false-negative rate:  $FNR = F^- / (F^- + T^-)$ ;  $F^+$  represents that the prediction is an

TABLE 1: Evaluation matrix.

		Prediction	
		Intrusion attack	Normal
Actual	Intrusion attack	$T^+$	$F^-$
	Normal	$F^+$	$T^-$

intrusion attack and the actual is normal; namely, prediction fails. Therefore, the number of samples of false positives in the system is denoted as the false-positive rate:  $FPR = F^+ / (F^+ + T^+)$ . The detection rate (DR) can be written as  $DR = T^+ / (F^- + T^+)$ .

**4.3. Experimental Dataset.** Recently, most of the standard examples used in the field of intrusion detection are still the KDDCUP99 dataset, and the test results in KDDCUP99 are better under certain conditions [32]. However, the KDDCUP99 dataset simulated 20 years ago is no longer suitable for the modern intelligent and complex attack methods, such as penetration utilization, SQL injection, APT, and complex hidden attack forms. The UNSW-NB15 dataset was created by the Australian Network Security Center (ANSC) in 2015 [33], which is a new dataset in the field of intrusion detection, which reflects the modern network traffic pattern. This dataset contains a large number of low occupancy intrusion and deep structured network traffic information and has 9 different types of modern attacks and 49 features. It has 5 attack types more than NSL-KDD, including 2540044 samples and 9 types of attacks, which are fuzzers, DoS, analysis, reconnaissance, exploit, shellcode, worm, backdoor, and generic. The UNSW-NB15 dataset contains 5.5 million records. Since the normal data sample in the UNSW-NB15 dataset is more than 10 times that of attack-type data sample, using a small number of sample oversampling will lead to excessive attack class duplicate samples, resulting in an overfitting phenomenon. If a simple undersampling strategy is used, normal samples will lose key information. Therefore, in order to reduce the error of experimental results caused by unbalanced data samples, the SMOTE oversampling method is adopted in this paper, as

TABLE 2: The description of the used dataset.

Type	Fuzzers	DoS	Analysis	Reconnaissance	Exploit	Shellcode	Worm	Backdoor	Generic	Normal
Training	17245	11357	2000	4000	33652	1285	125	1587	625	58220
Testing	6005	4072	655	15892	12854	320	42	550	127	32125

shown in formula (8). The principle is to use the existing attack sample data to find the random samples in the same kind of samples, and the linear difference will generate a new sample  $x$ , and repeat this process until the samples are balanced.

$$\hat{x} = x + \delta(x_i - x), \quad (8)$$

where  $\hat{x}$  is the new sample,  $x$  is the actual sample, and  $x_i$  is a randomly selected sample from a sample near  $x$ . The dataset is divided into two parts, including the training dataset and the test dataset, as shown in Table 2.

#### 4.4. Qualitative and Quantitative Analysis

**4.4.1. Effectiveness Analysis for the Spatiotemporal Module.** In order to verify the effectiveness of the introduced spatiotemporal module, different modules are tested under the condition that other parameter settings remain unchanged, as shown in Figure 4. It is not difficult to see from Figure 4 that the detection rate of the spatiotemporal module selected in this paper is higher than that of the other three modules, especially for the accuracy rate of normal data compared with the efficiency of other modules; the maximum difference is 4.30%, which can reflect the significant advantage of adding a spatial structure feature between classes in the spatiotemporal module.

**4.4.2. Effectiveness Analysis for LSTM Depth Level.** This paper analyzes the intrusion detection performance of the proposed convolutional neural network model and other four convolutional neural network structures and takes the detection accuracy and false alarm rate as the evaluation criteria of this experiment. The results are shown in Table 3. Through the comparative analysis, the detection accuracy of our proposed intrusion detection model is the highest, which is 1.53% higher than that of the IRES model [34]. Compared with the ResNet model with a low false alarm rate, its accuracy rate is increased by 5.82%; the false alarm rate is lower than that of the IRES model and LeNet model [35]. On the other hand, we can see that the depth of network structure is gradually deepened in the process of applying a convolutional neural network to intrusion detection, which is also a factor of a high false alarm rate of the comparison models. With the deepening of the network structure, the phenomenon of gradient vanish is obvious, but the special structure of our proposed convolutional neural network ensures that the problem of gradient vanish is improved in depth level.

**4.4.3. Performance Analysis.** In the experiment, the binary classification form is adopted, the input vector has 47 dimensional features, the intrusion attack vector is marked with 0,

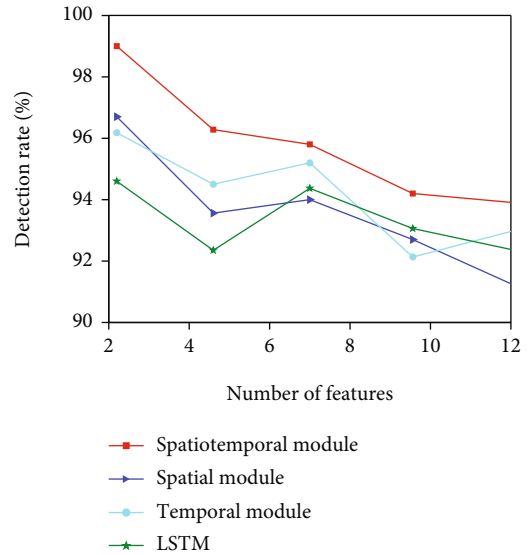


FIGURE 4: Detection rate for different modules.

TABLE 3: Comparison analysis.

Models	Accuracy (ACC)	DR	FNR	FPR
ResNet	91.36	1.35	0.25	0.27
LeNet	92.18	1.98	0.42	0.19
LSTM	97.22	2.25	0.23	0.22
IRES	96.25	3.59	0.28	0.25
Proposed	98.76	0.76	0.17	0.12

and the nonattack is marked with 1, which can improve the efficiency and the timeliness of the intrusion detection system. We can judge whether it is abnormal data through the model first and then classify the abnormal data by supervised learning. By using SMOTE sampling, it is divided into the training set and test set. There are 30 training datasets, with each dataset containing 3000 randomly selected samples, and 1500 test datasets, with each dataset containing 3000 randomly selected samples. In the process of feature extraction by a spatiotemporal network, the relationship between the number of iterations and the loss value is shown in Figure 5. After repeated iterations, the effect of the spatiotemporal model tends to be stable. The number of neurons in each layer was 47, 23, 23, and 23. In our improved LSTM model, three hidden layers are set up in LSTM. The number of neurons in the first layer is 128, that in the second layer is 256, and that in the third layer is 128. The batch size and epoch times are 200 and 800, respectively. When the learning

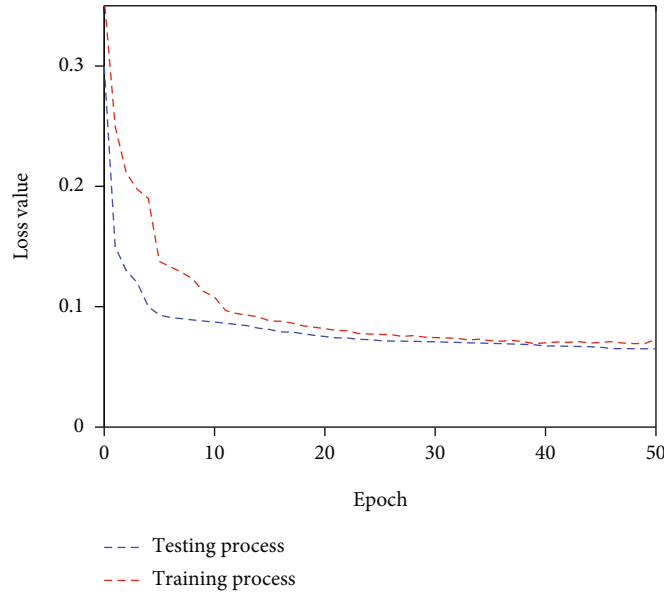


FIGURE 5: Relationship between the number of iterations and the loss value in training and testing processes.

rate of the model is too large or too small, the accuracy of the testing set is very low. Since the choice of the learning step has a great impact on the performance of the intrusion detection system, the time step is selected as 50 through repeated experiments.

In this experiment, `categorical_crossentropy`, which is specially used for solving multiclassification problems, is selected as the optimization objective function, and the Adam optimization algorithm is adopted to carry out the backpropagation training of the model. As shown in Figure 6, in order to highlight the ability of spatiotemporal feature extraction, the model in this paper is compared with LeNet [35], ResNet [36], and LSTM [37] without feature extraction. The results show that the accuracy of LeNet is not significantly improved in high-dimensional data samples, and the accuracy rate of LSTM is significantly lower than that of our proposed model. With the increase in the number of samples, the accuracy of LSTM shows a downward trend, which shows that the LSTM and spatiotemporal structure have strong data feature extraction ability, and with the increase in time complexity, the spatiotemporal structure has more and more advantages than the spatial structure [10, 11, 38]. Therefore, the detection effect of the intrusion detection algorithm in this paper is obviously better than that of the LeNet, ResNet, and LSTM intrusion detection model. The experiment selects the existing LSTM model, MLP (multilayer perception), ELM (extreme learning machine), and the popular deep belief model (DBN) as comparison models [12]. ResNet has multiple hidden layers and is fully connected with the input layer, which can be processed by nonlinear activation function [39]; DBN is a probability generation model, which has the characteristics of multiple hidden layer models and can learn the essence of the dataset by learning a deep nonlinear network structure [40-46]; ELM is a simple forward propagation by setting the connection weights and thresholds of the input layer and hidden layer

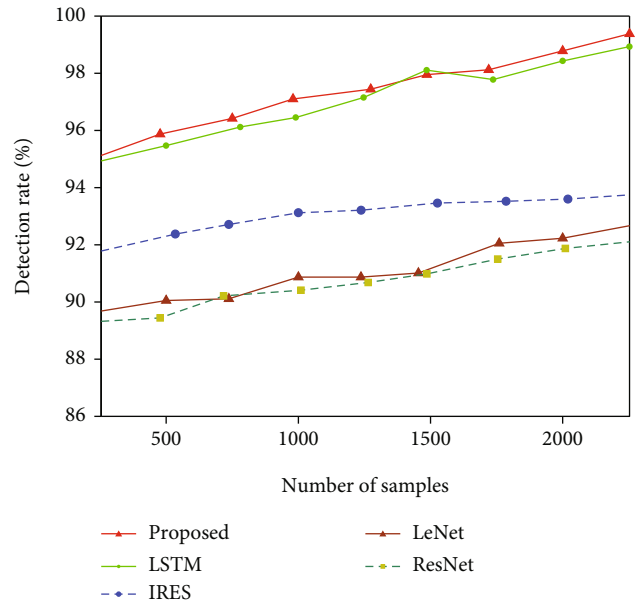


FIGURE 6: Comparison of detection rates for different models.

randomly and adjusting the weights without iteration, and the learning speed is improved. As can be seen from Table 3, the detection rate and false alarm rate of the proposed algorithm in this paper are improved compared with the current popular deep learning methods [47]. As shown in Figure 7, through the comparison of ROC curves, since the area of the ROC curve [48] of the proposed model is the largest, it can be proven that the proposed intrusion detection algorithm model not only improves the accuracy of intrusion detection but also significantly reduces the false alarm rate and achieves remarkable results in improving the model performance and detection efficiency.

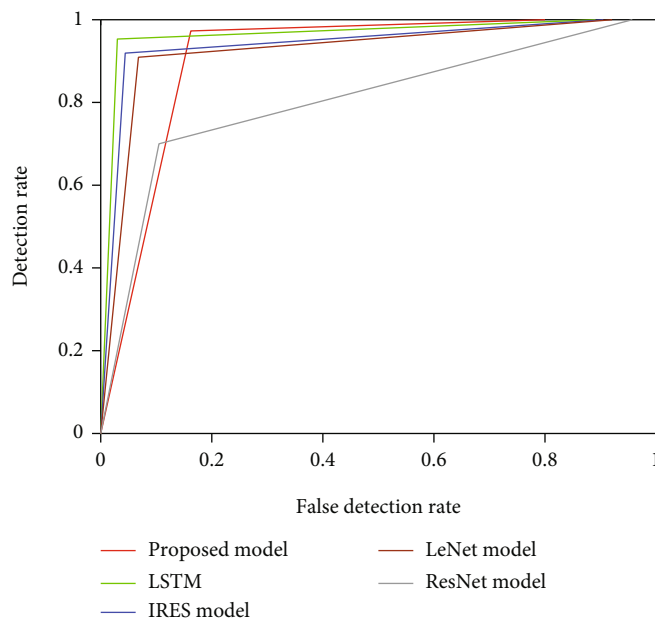


FIGURE 7: Comparison of ROC curves for different models.

## 5. Conclusion

Aimed at the existing problems in network intrusion detection, this paper proposes an improved LSTM combined with spatiotemporal structure for intrusion detection. The unsupervised spatiotemporal encoder is used to intelligently extract the spatial characteristics of network traffic data samples. It can not only retain the nonlocal characteristics of the data samples but also extract the deep features of the data samples. The model is trained and tested by using the NSL-KDD dataset which is more in line with the data characteristics of the new era than the traditional KDDCUP99 dataset. The experimental results show that the proposed intrusion detection model has achieved remarkable results in improving the accuracy, performance, and efficiency of intrusion detection. The model only works well in the simulation dataset but needs to be tested in the actual network environment to verify the real performance of the model, which will be our working direction in the future.

## Data Availability

The labeled dataset used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The author declares no conflicts of interest.

## Acknowledgments

This work was supported in part by the Cultural Science Research of Jiangsu Province under Grant 18YB27.

## References

- [1] M. Weizhi, "Intrusion detection in the era of IoT: building trust via traffic filtering and sampling," *Computer*, vol. 51, no. 7, pp. 36–43, 2018.
- [2] Z. Sun, Y. Xu, G. Liang, and Z. Zhou, "An intrusion detection model for wireless sensor networks with an improved V-detector algorithm," *IEEE sensors journal*, vol. 18, no. 5, pp. 1971–1984, 2018.
- [3] S. Manimurugan, "Intrusion detection in cloud environment using hybrid genetic algorithm and back propagation neural network," *International Journal of Communication Systems*, vol. 56, no. 9, pp. 258–267, 2018.
- [4] L. Li, H. Zhang, H. Peng, and Y. Yang, "Nearest neighbors based density peaks approach to intrusion detection," *Chaos Solitons & Fractals*, vol. 110, pp. 33–40, 2018.
- [5] A. Karami, "An anomaly-based intrusion detection system in presence of benign outliers with visualization capabilities," *Expert Systems with Applications*, vol. 108, no. 6, pp. 36–60, 2018.
- [6] H. Jun-Ho, "Implementation of lightweight intrusion detection model for security of smart green house and vertical farm," *International Journal of Distributed Sensor Networks*, vol. 14, no. 4, 2018.
- [7] X. An, X. Zhou, X. Lü, F. Lin, and L. Yang, "Sample selected extreme learning machine based intrusion detection in fog computing and MEC," *Wireless Communications and Mobile Computing*, vol. 2018, 10 pages, 2018.
- [8] R. Abdulhammed, M. Faezipour, A. Abuzneid, and A. AbuMallouh, "Deep and machine learning approaches for anomaly-based intrusion detection of imbalanced network traffic," *Electronics Letters*, vol. 56, no. 19, pp. 23–39, 2018.
- [9] G. Mamalakis, C. Diou, A. L. Symeonidis, and L. Georgiadis, "Of daemons and men: reducing false positive rate in intrusion detection systems with file system footprint analysis," *Neural Computing & Applications*, vol. 31, no. 3, pp. 7755–7767, 2018.

- [10] Y. Jiang, Y. Zhang, C. Lin, D. Wu, and C.-T. Lin, "EEG-based driver drowsiness estimation using an online multi-view and transfer TSK fuzzy system," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 3, pp. 1752–1764, 2021.
- [11] Y. Jiang, G. Xiaoqing, D. Wu et al., "A novel negative-transfer-resistant fuzzy clustering model with a shared cross-domain transfer latent space and its application to brain CT image segmentation," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 18, no. 1, pp. 40–52, 2021.
- [12] Y. Jiang, K. Zhao, K. Xia et al., "A novel distributed multitask fuzzy clustering algorithm for automatic MR brain image segmentation," *Journal Medical Systems*, vol. 43, no. 5, 2019.
- [13] K. Leyli, S. Erkay, and A. Halit, "Intrusion detection over encrypted network data," *The Computer Journal*, vol. 63, no. 4, pp. 604–619, 2020.
- [14] D. He, Q. Qiao, Y. Gao et al., "Intrusion detection based on stacked autoencoder for connected healthcare systems," *IEEE Network*, vol. 33, no. 6, pp. 64–69, 2019.
- [15] M. F. Umer, M. Sher, and Y. Bi, "A two-stage flow-based intrusion detection model for next-generation networks," *Plos One*, vol. 13, no. 1, 2018.
- [16] X. Tan, S. Su, Z. Huang et al., "Wireless sensor networks intrusion detection based on SMOTE and the random forest algorithm," *Sensors*, vol. 133, no. 26, pp. 64–69, 2019.
- [17] J. Liang, M. Ma, M. Sadiq, and K.-H. Yeung, "A filter model for intrusion detection system in vehicle ad hoc networks: a hidden Markov methodology," *Knowledge-Based Systems*, vol. 23, no. 16, pp. 214–219, 2020.
- [18] M. K. Prasath and B. Perumal, "A meta-heuristic Bayesian network classification for intrusion detection," *International Journal of Network Management*, vol. 29, no. 3, pp. e2047.1–e2047.12, 2019.
- [19] Z. Liu, M. Zhou, W. Nie, L. Xie, and Z. Tian, "Indoor intrusion detection based on fuzzy membership-aided Dempster-Shaper theory," *Journal of Intelligent and Fuzzy Systems*, vol. 38, no. 4, pp. 3687–3696, 2020.
- [20] G. Yang, X. Yu, L. Xu, Y. Xin, and X. Fang, "An intrusion detection algorithm for sensor network based on normalized cut spectral clustering," *PLoS ONE*, vol. 14, no. 10, 2019.
- [21] Z. Tang, S. Wang, J. Huo, H. Guo, H. Zhao, and Y. Mei, "Bayesian framework with non-local and low-rank constraint for image reconstruction," *Journal of Physics Conference Series*, vol. 787, 2017.
- [22] D. B. Gothawal and S. V. Nagaraj, "Anomaly-based intrusion detection system in RPL by applying stochastic and evolutionary game models over IoT environment," *Wireless Personal Communications*, vol. 110, no. 3, pp. 789–795, 2019.
- [23] H. Deng, "An improved two-steps saliency detection algorithm based on binarized normed gradients and nuclear norm model in video sequences," *Journal of Information Hiding and Multimedia Signal Processing*, vol. 9, no. 4, pp. 841–852, 2018.
- [24] S. Priyanga, M. R. G. Raman, S. S. Jagtap, N. Aswin, K. Kirthivasan, and V. S. S. Sriram, "An improved rough set theory based feature selection approach for intrusion detection in SCADA systems," *Journal of Intelligent and Fuzzy Systems*, vol. 36, no. 5, pp. 3993–4003, 2019.
- [25] K. R. C. Boni, L. Xu, Z. Chen, and T. D. Baddoo, "A security concept based on scaler distribution of a novel intrusion detection device for wireless sensor networks in a smart environment," *Sensors*, vol. 20, no. 17, 2020.
- [26] X. Tian, H. Li, and H. Deng, "Object tracking algorithm based on improved Siamese convolutional networks combined with deep contour extraction and object detection under airborne platform," *Journal of Imaging Science and Technology*, vol. 12, no. 6, pp. 2369–2375, 2020.
- [27] B. S. Bhati, G. Chugh, F. Al-Turjman, and N. S. Bhati, "An improved ensemble based intrusion detection technique using XGBoost," *Transactions on Emerging Telecommunications Technologies*, vol. 18, no. 11, pp. 3909–3917, 2020.
- [28] X. Tian, H. Li, and H. Deng, "Object tracking algorithm based on improved context model in combination with detection mechanism for suspected objects," *Multimedia Tools and Applications*, vol. 78, no. 4, pp. 259–268, 2019.
- [29] G. Spathoulas, G. Theodoridis, and G. P. Damiris, "Using homomorphic encryption for privacy-preserving clustering of intrusion detection alerts," *International Journal of Information Security*, vol. 62, no. 33, 2020.
- [30] G. Folino, F. S. Pisani, and L. Pontieri, "A GP-based ensemble classification framework for time-changing streams of intrusion detection data," *Soft Computing*, vol. 24, no. 23, pp. 17541–17560, 2020.
- [31] K. Vieira, F. L. Koch, J. B. M. Sobral, C. B. Westphall, and J. L. de Souza Leao, "Autonomic intrusion detection and response using big data," *IEEE Systems Journal*, vol. 14, no. 2, pp. 1984–1991, 2019.
- [32] D. Zheng, Z. Hong, N. Wang, and P. Chen, "An improved LDA-based ELM classification for intrusion detection algorithm in IoT application," *Sensors*, vol. 20, no. 6, 2020.
- [33] H. Zhang, L. Huang, C. Q. Wu, and Z. Li, "An effective convolutional neural network based on SMOTE and Gaussian mixture model for intrusion detection in imbalanced dataset," *Computer Networks*, vol. 177, 2020.
- [34] R. H. Dong, X.-Y. Li, Q.-Y. Zhang, and H. Yuan, "Network intrusion detection model based on multivariate correlation analysis – long short-time memory network," *IET Information Security*, vol. 14, no. 2, pp. 166–174, 2020.
- [35] V. Dutta, M. Choraś, M. Pawlicki, and R. Kozik, "A deep learning ensemble for network anomaly and cyber-attack detection," *Sensors*, vol. 20, no. 16, 2020.
- [36] N. Ding, H. X. Ma, H. Gao, Y. H. Ma, and G. Z. Tan, "Real-time anomaly detection based on long short-term memory and Gaussian mixture model," *Computers & Electrical Engineering*, vol. 79, no. 5, 2019.
- [37] A. Diro and N. Chilamkurti, "Leveraging LSTM networks for attack detection in fog-to-things communications," *IEEE Communications Magazine*, vol. 56, no. 9, pp. 124–130, 2018.
- [38] L. Nicholas, S. Y. Ooi, Y. H. Pang, S. O. Hwang, and S.-Y. Tan, "Study of long short-term memory in flow-based network intrusion detection system," *Journal of Intelligent and Fuzzy Systems*, vol. 35, pp. 5947–5957, 2018.

## Research Article

# Video Stream Session Migration Method Using Deep Reinforcement Learning in Cloud Computing Environment

Lingling Li  and Huixia Liu 

*Department of Cyber Security, Henan Police College, Zhengzhou, Henan 450000, China*

Correspondence should be addressed to Lingling Li; [lll007@hnp.edu.cn](mailto:lll007@hnp.edu.cn)

Received 28 January 2021; Revised 11 March 2021; Accepted 26 March 2021; Published 12 April 2021

Academic Editor: Shan Zhong

Copyright © 2021 Lingling Li and Huixia Liu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In the resource scheduling of streaming Media Edge Cloud (MEC), in order to balance the cost and load of migration, this paper proposes a video stream session migration method based on deep reinforcement learning in cloud computing environment. First, combined with the current popular OpenFlow technology, a novel MEC architecture is designed, which separates streaming media service processing in application layer from forwarding path optimization in network layer. Second, taking the state information of the system as the attribute feature, the session migration is calculated, and gradient reinforcement learning is combined with in-depth learning and deterministic strategy for video stream session migration to solve the user request access problem. The experimental results show that the method has a better request access effect, can effectively improve the request acceptance rate, and can reduce the migration cost, while shortening the running time.

## 1. Introduction

In recent years, with the maturity of cloud computing technology, streaming media services are gradually transforming to cloud form, that is, streaming Media Cloud. Streaming media cloud pushes the content requested by users to the edge of the network by placing media edge cloud in different geographical locations, so that to reduce the user response delay and reduce the traffic load of the main network [1]. At the same time, the subcloud can adapt to the changes of system load and the size of user requests, so that to effectively solve the problem of traditional streaming media services [2].

In streaming Media Edge Cloud (MEC), system resources are virtualized into resource pools to ensure service transparency. Cloud resource allocation is automatically adjusted by the cloud platform according to the scale of actual demand, so how to allocate system resources in real-time to meet user needs. Under the condition of limited resource allocation, the fluctuation and randomness of user request mode will make the system load unbalanced and affect the access effect of user request [3, 4].

In order to solve the above problems, domestic and foreign scholars have proposed a migration-based task scheduling method for streaming media. Ref. [5] proposed a session migration strategy based on dynamic threshold allocation (SMS-DTA). According to the popularity distribution, the session allocation thresholds of all kinds of videos on each server are determined, and the user request access is guided by the allocation thresholds. Ref. [6] proposed resource dispatch based on data priority (RDDP) algorithm. However, considering the impact of the urgency and scarcity of data blocks on priority, quantitative calculation is not given. Only the balance factor is used to measure the quantitative relationship between them, and the influence of time factor on emergency quantification is omitted. Ref. [7] proposed a direct access storage device (DASD) hopping algorithm to migrate sessions of nodes with different loads in order to maintain the load balance of hard disk. However, due to the lack of self-adaptability, it is difficult to adjust the strategy according to the system operation scenario. Moreover, the mathematical model is relatively complex and the calculation is large, which cannot solve the problem of large-scale resource allocation.

Ref. [8] explored how to make the streaming media edge cloud admit more requests via online session migration and proposed an adaptive strategy of online session migration. Besides the load information, the video popularity is adopted for obtaining the allocation thresholds of different videos on each server, and a new request would be admitted under the guidance of the obtained threshold distribution. Specially, when the video popularity varies, the allocation thresholds would be recalculated. Ref. [9] proposed a joint optimization algorithm of session migration and video deployment, the proposed strategy is more adaptive to dynamic fluctuation of video popularity, and thus gains a flexible balance between service cost and quality. The trace-driven experiment verified the effectiveness of the proposed method.

According to the resource allocation of streaming media edge cloud, in order to balance the cost and load of migration, considering the cost of migration, load balancing, and other constraints, this paper proposes a video stream session migration method based on deep reinforcement learning. Based on the current popular OpenFlow technology, a novel MEC architecture is designed, which separates streaming media service processing in application layer from forwarding path optimization in network layer to ensure service transparency. The main innovations are as follows:

- (1) This paper improves the resource utilization by effectively utilizing the state information of the MEC system, combining in-depth learning and deterministic strategy for video stream session migration
- (2) This paper proposes a session migration computing model to process user requests more scientifically, maximize the access rate of user requests, and control the migration cost appropriately, at the same time, make the system achieve load balancing as far as possible

## 2. Streaming Media Edge Cloud Architecture

Streaming Media Edge Cloud is located on the edge of the network, which is responsible for local video services. As shown in Figure 1, combined with the current popular OpenFlow technology, this paper designs a novel MEC architecture. The whole MEC is composed of streaming media server, business management server, and OpenFlow controller and switch, in which the streaming media server is responsible for providing media streaming to users; the business management server is mainly responsible for the access scheduling of user requests, generating migration strategies and sending them to OpenFlow controller; the OpenFlow controller and switch, on the one hand, it constitutes a media stream distribution network, on the other hand, it is responsible for the actual implementation of session migration; OpenFlow controller is responsible for generating flow tables according to migration strategy and sending them to switches; OpenFlow switch completes the modification and forwarding of data packets according to flow tables.

By introducing MEC architecture, streaming media service processing in application layer is separated from for-

warding path optimization in network layer, and transparency of video service is realized.

## 3. Session Scheduling Strategy Based on Deep Reinforcement Learning

Assuming that the video content provided by the MEC system has  $I$  kind of video content, the  $i$  kind of video is represented by  $v_i$ . Each kind of video is encoded at a constant bit rate and serves at the same bit rate [10, 11]. Assuming that the total number of MEC streaming media servers is  $J$ , the  $j$  server is represented by  $M_j$ , and  $D$  is defined as the video deployment matrix with the size of  $I \times J$ , and the element  $d_{ij} \in \{0, 1\}$  represents whether or not a copy of  $v_i$  is deployed on  $M_j$ . Assuming that all servers are homogeneous, and a single server can provide up to  $C$  streaming sessions at the same time, as well as up to  $S$  videos [12].

$K$  is defined as a session distribution matrix with the size of  $I \times J$ . A single element  $k_{ij} \in [0, 1]$  represents the ratio of all sessions of video  $v_i$  on server  $M_j$  to the total service capacity (JC) of the system.  $G$  is defined as a server adjacency matrix with the size of  $J \times J$ . Element  $g_{nj} \in \{0, 1\}$  denotes whether there are sessions on  $M_n$  server that can be migrated to the server, where  $n = 1, 2, \dots, J$ .

Define  $l_j$  as the load of streaming media server  $M_j$ , that is, the total number of access sessions, then:

$$l_j = \sum_{i=1}^I (k_{ij} \times JC). \quad (1)$$

Define  $\bar{L}$  as the average load of all streaming media servers in the system, then:

$$\bar{L} = \sum_{j=1}^J l_j / N. \quad (2)$$

In this paper, the state information of the MEC system is taken as an attribute feature, and the decision-maker and value function are fitted by deep convolution neural network combined with reinforcement learning elements such as state space, action set, and return function. In order to improve the efficiency of the algorithm, the deterministic strategy gradient is used to train the neural network.

**3.1. Session Scheduling Model.** For streaming media edge cloud system, the goal of reinforcement learning is to access the video request to the most suitable server independently according to the current MEC system status and video request according to the experience strategy. Then, according to the load state of the server, the optimal migration method for the current incoming user requests is obtained by using the migration video strategy to perform the request access or one-step session migration action [13, 14].

In this paper, the deep reinforcement learning method is applied to session scheduling in streaming media edge cloud, and its session migration method is shown in Figure 2.

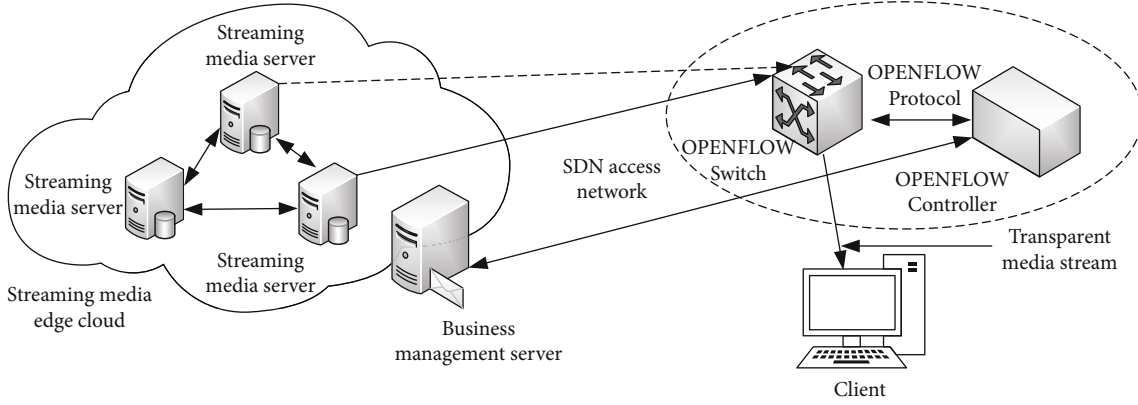


FIGURE 1: The novel architecture of MEC.

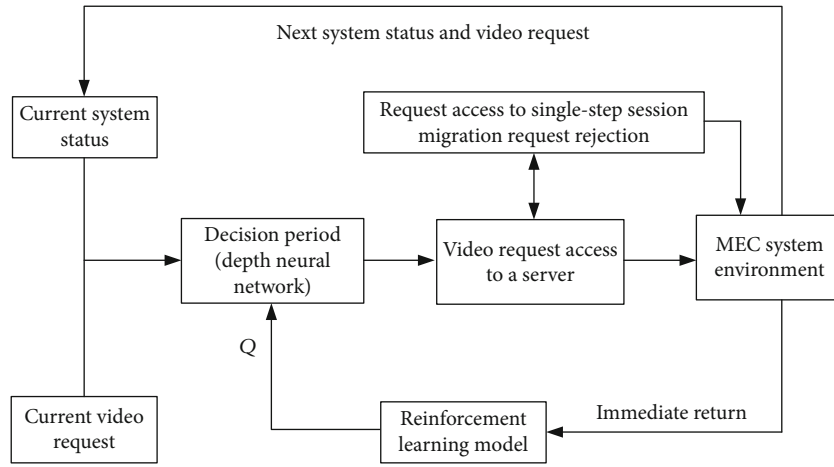


FIGURE 2: Conversation transfer model based on deep reinforcement learning.

In Figure 2, for the current step video request  $v_i$ , the decision-making action of the decision-maker is to connect the video request to a server, assuming that the server accessed is  $M_j$ , then the strategy of moving out video is: if  $M_j$  is not full, no video needs to be moved out, set the number of moving out video to be 0, corresponding to the request access; if  $M_j$  is full, it needs to move out the video and move out the video. The set of numbers is  $\{v_m | k_{mj} \geq 1, m \neq i, m \leq I\}$ , corresponding to one-step session migration.

**3.2. Enhanced Learning Model of Conversation Transfer.** According to the characteristics of the problem, the state of  $t$  time step in the MEC system is defined as follows:

$$s_t = [G_t : R_t : D_t : K_t]. \quad (3)$$

Among them,  $G_t$  is the server adjacency matrix and the size is  $J \times J$ , which indicates whether session migration can be carried out among servers,  $R_t$  is the video request matrix of  $t$  time step, the size is  $I \times J$ , the elements of one row in the video request matrix are all 1, the other elements are all 0, and the corresponding video number of the row with the elements all 1 is the appropriate one. The former video request  $D_t$  is the video deployment matrix and the size is  $I$

$\times J$ , which reflects the deployment of video copies in the MEC system.  $K_t$  is the session distribution matrix and the size is  $I \times J$ , which will reflect the distribution of video sessions in the MEC system. Every time a new video request is processed, the system will undergo a state transition [15–17].

Since the task is to decide which server to access or reject the request based on the current MEC system status and video request, the action is defined as the server number  $M_j$  to which the video request is accessed, where  $i = 1, 2, \dots, I$ . For the current step video request  $v_i$ , the optional action set is shown in Formula (4). When  $v_i$  accesses MEC directly or through session migration, the set of optional actions is the set of servers deployed with video  $v_i$ ; when rejecting video request  $v_i$ , the corresponding action is 0.

$$a_t = \begin{cases} M_j | d_{ij} = 1, j \leq J, \text{insert,} \\ 0, \text{refuse.} \end{cases} \quad (4)$$

If video request  $v_i$  is accessed to server  $M_j$  according to the decision-making action, this paper chooses the deployment of video  $v_i$  on server  $M_j$ , the load of server  $M_j$ , and the variance of load balance of MEC system after executing the action as the immediate return function. The video



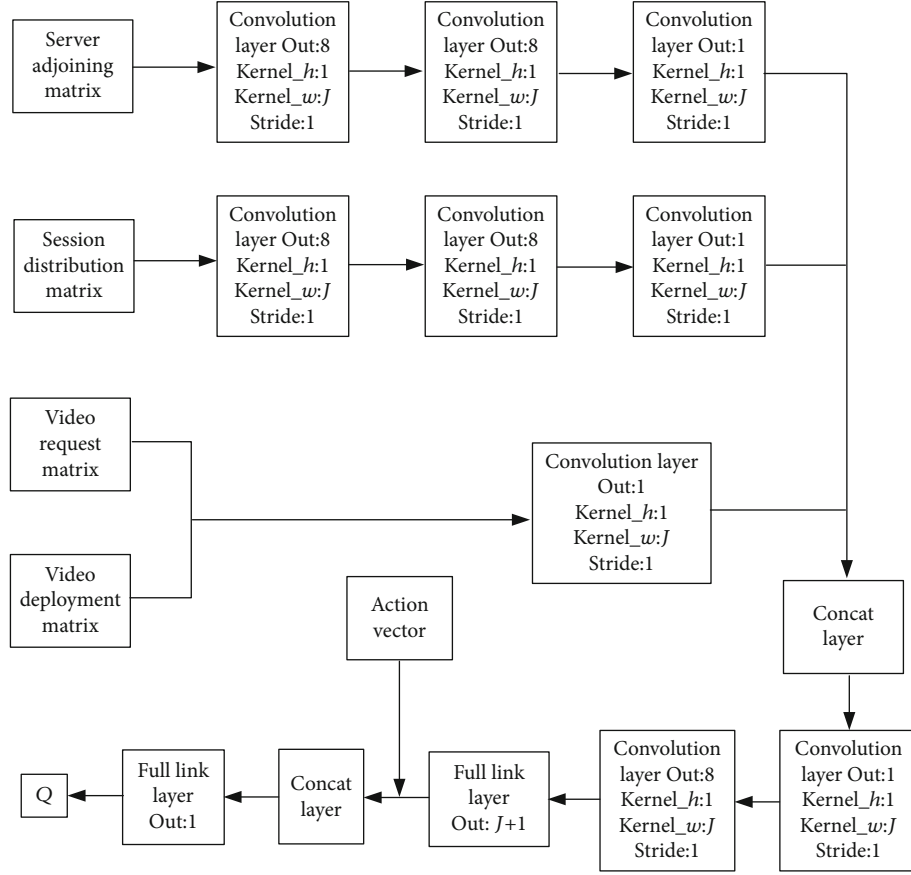


FIGURE 3: Q network model based on bible network.

deployment on the server, the load value of the server, and the load balancing variance of the system are different, so the load balancing variance of the system is normalized. The load balancing variance function is defined as:

$$\sigma = \arctan \left( \frac{\sum_{j=1}^J (l_j - \bar{L})^2}{J} \right) \times \frac{2}{\pi}. \quad (5)$$

Since the variance of load balancing is the inverse of variance, the formula above shows that the larger the variance  $\sigma$  of load balancing, the more balanced the load of the system.

If the video requested in time step  $t$  is  $v_i$ , the quotation value returned by action  $a_t$  of that step is

$$r(s_t, a_t) = \begin{cases} \omega_1 \times d_{ij} + \omega_2 \times \left(1 - \frac{l_j}{C}\right) + \omega_3 \times \sigma, & \text{insert,} \\ -1 + \omega_1 \times d_{ij} + \omega_3 \times \sigma, & \text{transfer,} \\ -1, & \text{refuse,} \end{cases} \quad (6)$$

where  $i = 1, 2, \dots, I$ ,  $j = 1, 2, \dots, J$ , when video  $v_i$  is deployed on the server  $M_j$  corresponding to decision action  $a_t$ , there will be corresponding reward value  $\omega_1$ . If video  $v_i$

is not deployed on server  $M_j$ , the action is not a reasonable access action. It is not in the optional action set, reward value 0, and reward value  $1 - l_j/C$  represents the remaining service capability of the server. When the server is full, that is, the residual service capacity is 0, the reward value is 0. When migration occurs, because session migration has a certain cost, the reward value is reduced by 1 as the corresponding penalty. When the action is to reject the video request, the return value is set to -1.  $\omega_1$ ,  $\omega_2$ , and  $\omega_3$  represent the weights of the returns from the three optimization objectives, respectively. The weights can be set according to the importance of the optimization objectives, but the sum of the three weights must satisfy  $\sum_{i=1}^3 \omega_i = 1$ .

Defined in MEC system state  $s_t$ , after taking action  $a_t$ , if strategy  $\mu$  is continuously implemented, the expected value of immediate return is action-value function. Defined Bellman equation as follows:

$$Q^\mu(s_t, a_t) = E[r(s_t, a_t) + \gamma Q^\mu(s_{t+1}, \mu(s_{t+1}))], \quad (7)$$

where  $r(s_t, a_t)$  is the immediate return value after taking action  $a_t$  under the state  $s_t$  of the MEC system. In the whole session scheduling process, the above equation is the final solution of the equation, and the optimal scheduling strategy is obtained by solving the equation.

**3.3. Migration Computing Model.** In MEC architecture, the migration cost can be expressed as  $\sum_{i=1}^I \sum_{j=1}^J \max(k_{ij} - a_{ij}, 0)$  by the number of migrated sessions. In addition, this paper specifies the maximum threshold  $M$ , i.e.,  $\sum_{i=1}^I \sum_{j=1}^J \max(k_{ij} - a_{ij}, 0) \leq M$ , for a single migration cost, where the value of  $M$  is determined by OpenFlow's flow processing capability.

Under the unbalanced load distribution, the full-loaded servers can continue to access new requests only if some sessions are moved out. Therefore, whether the load is balanced or not will indirectly affect the cost of migration. In practice, due to the fluctuation of request distribution, all kinds of video requests do not arrive strictly according to popularity. The scheme of optimizing the acceptance rate mentioned above can easily lead to an unbalanced load and increase the cost. Therefore, a goal of load balancing maintenance is introduced.

- (1) For new requests arriving immediately  $r$ , due to the directive allocation threshold,  $r$  can only be connected to nodes that have not yet reached  $M$ , in order to minimize the load imbalance. Therefore, the following new optimization objectives have been added

$$\min \left\{ \sum_{i=1}^I \sum_{j=1}^J \max(a_{ij}^r - a_{ij}, 0) \right\}, \quad (8)$$

where  $A^r$  is a constant matrix, the calculation method is: for  $v_r$ , assuming that  $M_j$  is the node deployed  $v_r$  with the smallest load, the corresponding element  $a_{rj}^r = k_{rj} + 1$ ; of course, the corresponding element  $a_{rj}^r = k_{rj} + 1, i \neq r$  on the other nodes. In addition, for the rest of the video  $v_i, i \neq r$ , the corresponding element is  $a_{rj}^r = k_{rj}, \forall j \leq 1$ .

- (2) For all subsequent arrival requests, in order to connect them to the minimum load node, it is necessary to ensure that each allocation threshold is larger than the number of sessions [18]. In addition, considering the continuity and randomness of request arrival, the difference between the allocation threshold and the number of sessions should also be related to the arrival of requests and other factors [19, 20]. Therefore, the following new optimization objectives have been added

$$\min \left\{ \sum_{i=1}^I \sum_{j=1}^J \theta_i \times \max(a_{ij}^* - a_{ij}, 0) \times \gamma_j \right\}, \quad (9)$$

where  $A^*$  is a constant matrix, assuming that  $R_i$  represents the number of requests that  $v_i$  has not yet arrived, it can be approximately expressed as  $R_i = \max(JCp_i - \sum_{j=1}^J k_{ij}, 0)$ . For the node set  $N_i = \{j \mid d_{ij} = 1, j \leq J\}$  of deployment  $v_i$ , the corresponding element  $a_{ij}^* = k_{ij} + R_i / \sum_{j=1}^J d_{ij}$ , and for the remaining nodes, the corresponding element  $a_{ij}^* = 0$ .

$\theta_{1 \times I}$  and  $\gamma_{J \times 1}$  are weight vectors, considering that popular video is more likely to affect the load distribution, the weight  $\theta_i$  is desirable  $p_i$ ; considering that lightweight nodes should allocate larger thresholds, the weight  $\gamma_i$  is desirable  $(1 - l_j/C)$ .

From the effect point of view, the smaller the load of nodes, the larger the allocation threshold to undertake more requests for access. However, since this optimization strategy is adopted after the start of MEC, the load of each node is basically balanced, so the abovementioned average allocation processing can still achieve the desired effect.

In addition to the limitation of the cost of a single migration, the following constraints should be considered: the service capacity limitation of each server; the value range limitation of  $a_{ij}$ ; and the principle of "no reduction in the number of actual sessions."

In summary, with session assignment matrix  $A$  as a decision variable, the migration computation model can be expressed as follows:

$$\begin{aligned} \text{Obj 1 : } & \min \left\{ \sum_{i=1}^I \sum_{j=1}^J \max(a_{ij}^r - a_{ij}, 0) \right\}, \\ \text{Obj 2 : } & \min \left\{ \sum_{i=1}^I \sum_{j=1}^J \theta_i \times \max(a_{ij}^* - a_{ij}, 0) \times \gamma_j \right\}, \end{aligned} \quad (10)$$

where  $A$  and  $A^*$  are const matrix.  
Subject to:

$$\begin{aligned} & \sum_{i=1}^I \sum_{j=1}^J \max(k_{ij} - a_{ij}, 0) \leq M, \\ & \sum_{j=1}^J a_{ij} \geq \sum_{j=1}^J k_{ij}, \forall i \leq I, \\ & \sum_{j=1}^J a_{ij} \geq \alpha p_i, \forall i \leq I, \\ & \sum_{i=1}^I a_{ij} \geq 1/J, \forall j \leq J, \\ & 0 \leq a_{ij} \leq d_{ij}, \forall i \leq I, \forall j \leq J. \end{aligned} \quad (11)$$

### 3.4. Scheduling Algorithm Based on Reinforcement Learning

**3.4.1. Choice of Behavior Strategies.** This paper chooses deterministic behavior strategies and defines a function  $\mu$ , which is expressed as:

$$a_t = \mu(s_t). \quad (12)$$

The behavior of each step can be obtained by calculating function  $\mu$ . Function  $\mu$  is simulated by using convolutional neural network. The network is a strategy network with a parameter of  $\theta^\mu$ . A function  $J(\mu)$  is used to measure the

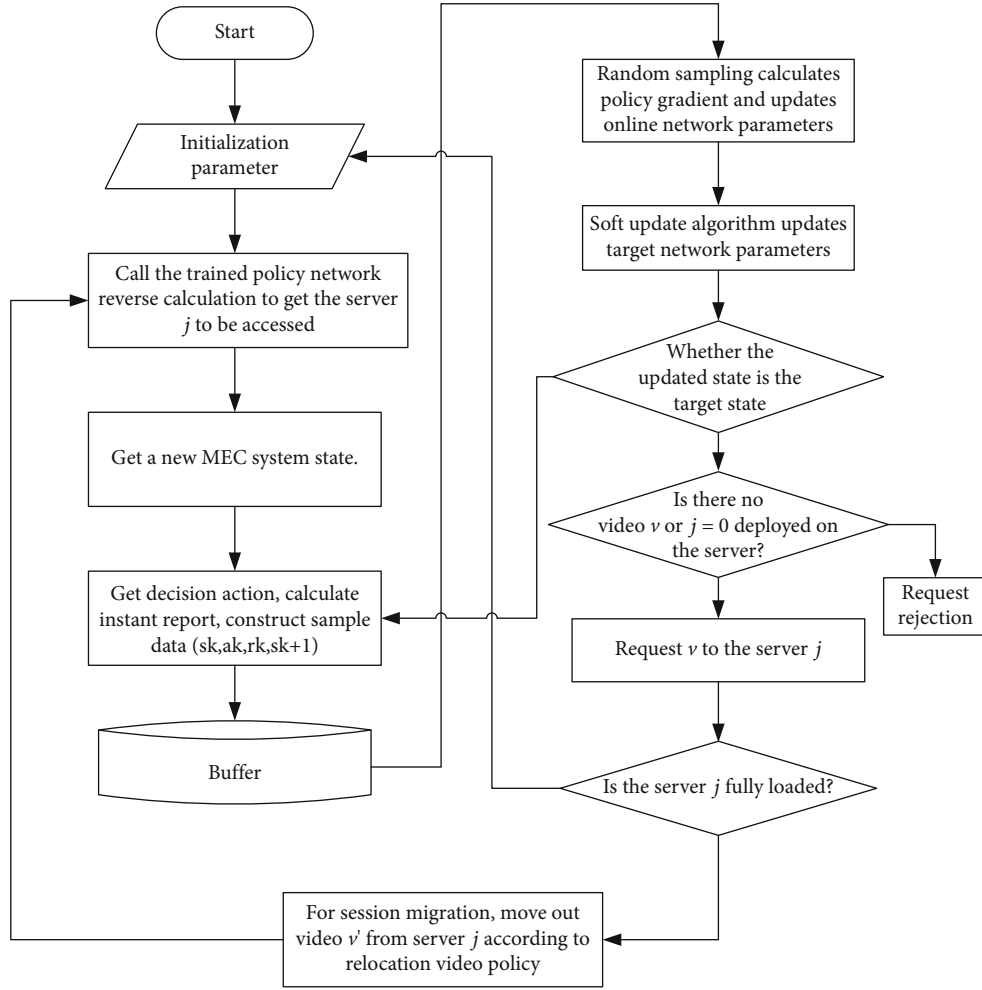


FIGURE 4: Strengthened learning algorithms for video session migration flow chart.

performance of strategy  $\mu$ , which is defined as:

$$J(\mu) = E[Q^\mu(s_t, \mu(s_t))], \quad (13)$$

where  $s_t$  is the state of the system,  $Q^\mu(s_t, \mu(s_t))$  is in each state, if the action  $a_t$  is selected according to policy  $\mu$ , the  $Q$  value can be generated, that is,  $J(\mu)$  is the expected value of  $Q^\mu(s_t, \mu(s_t))$  when policy  $\mu$ . Therefore, the optimal behavior strategy is the strategy  $\mu$  which maximizes  $J(\mu)$ , that is,

$$\mu = \arg \max (J(\mu)). \quad (14)$$

Network input is MEC system state, that is, video request matrix, video deployment matrix, session distribution matrix, and server adjacency matrix with size  $J \times J$ . The eigenvectors of the video request matrix and the video deployment matrix represent the deployment information of the current video on the server. The size of the eigenvectors is  $I \times J$ . The three eigenvectors are connected through concat layer. Finally, the probability distribution of server number is obtained by using Softmax classifier. The dimension is  $d$ , and the decision-making action is the server number corresponding to the maximum probability.

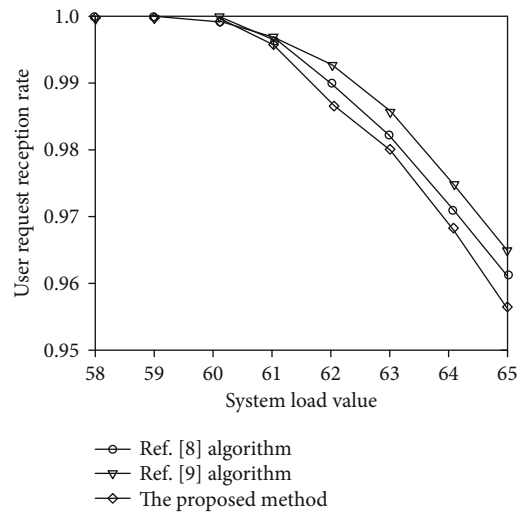


FIGURE 5: The relation between the receiving rate of user request and system load under Zipf distribution.

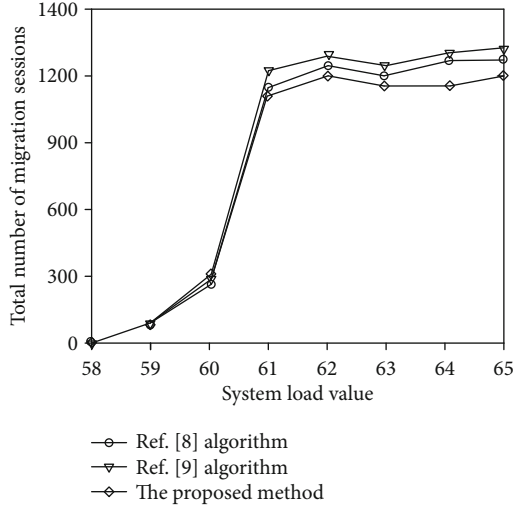


FIGURE 6: The relationship between total migration sessions and system load under Zipf distribution.

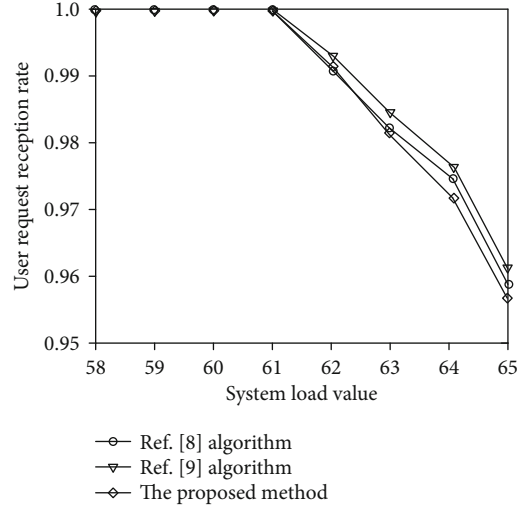


FIGURE 8: The relation between the receiving rate of user request and system load under random uniform distribution.

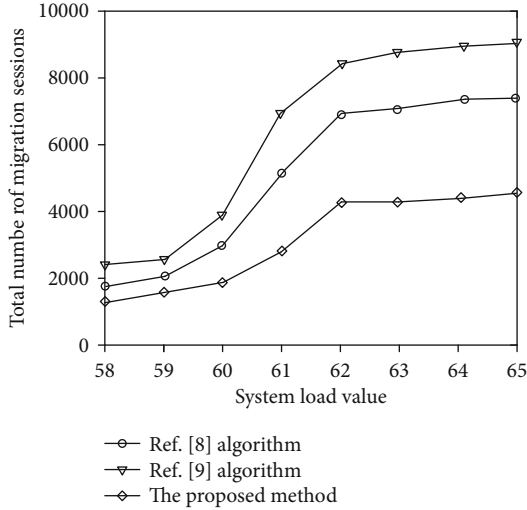


FIGURE 7: The relation between running time and system load under Zipf distribution.

In order to make it more exploratory, on the basis of the deterministic strategy, behavior search is added, that is, 30% of the actions are randomly selected in the optional action space, and the remaining actions are the output of the strategy network.

**3.4.2. Iterative Value Calculation.** In this paper, the convolution neural network is used to simulate  $Q$  function. The network is called  $Q$  network. Its parameter is  $\theta^Q$ . The model of  $Q$  network is shown in Figure 3.

The input of  $Q$  network is the MEC system state and action vector, and the action vector is the result of transforming the probability distribution vector of the output of policy network into one-hot vector, the size of which is  $I \times J$ . In network training, input sample data is highly correlated with time, and direct training is not easy to converge. In order to

break the correlation between data, the method of “experience playback” is used to save the generated sample data into the buffer, and the sample data used in training is randomly extracted from the buffer.

In the process of network training, this paper uses the target network method to establish the copy  $\theta^\mu$  and  $\theta^Q$  of the policy network and  $Q$  network to calculate the target value, and then update the original network slowly in the proportion of  $\tau$ . Through this network learning method, the learning process will be more stable and convergence will be more guaranteed. The flow chart of reinforcement learning algorithm based on deterministic strategy gradient is shown in Figure 4.

## 4. Experiment and Analysis

**4.1. Parameter Setting.** In the environment of the MEC system simulation, the environment parameters are as follows: the total number of streaming media servers  $J = 30$ , the capacity of each streaming media server  $S = 40$ , the maximum number of service sessions  $C = 100$ , and the number of video types  $I = 350$ . At the same time, assuming that the arrival rate of users’ requests obeys Poisson distribution of  $\zeta$  requests per minute, the value of  $\zeta$  ranges from 58 to 65. The average playback time is set to 30 minutes, and the system can support 2000 ( $J \times C$ ) user video requests concurrently in one playback time. Therefore, when  $\zeta = 65(1900/30)$ , the system reaches full load. The video content requested by users obeys Zipf distribution and random uniform distribution, respectively.

In order to analyze the effectiveness and practicability of deep reinforcement learning algorithm, this paper programmed on tensorflow platform and applied it in the MEC system session scheduling strategy. The parameters of the algorithm are as follows: the learning rate of policy network is 0.0001, the learning rate of  $Q$  network is 0.001, the discount coefficient is 0.95, the capacity of buffer is 1,00000, the preheating coefficient of buffer is 1,000, the number of iterations is 100,000, the upper limit of time step is 60 steps,

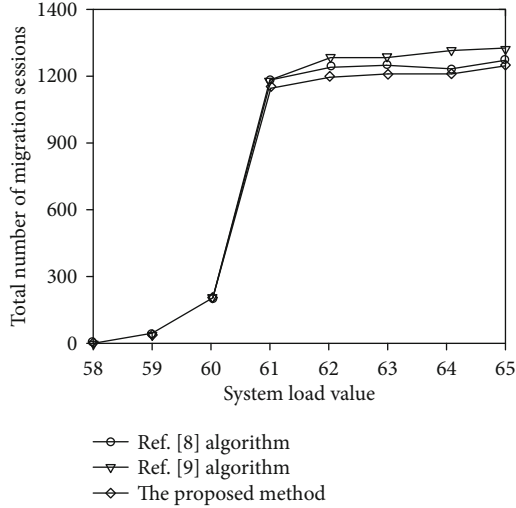


FIGURE 9: The relationship between total migration session number and system load under random uniform distribution.

the number of samples for each iteration is 600, and the weight coefficient in the return function is  $\omega_1 = 0.8$ ,  $\omega_2 = 0.1$ ,  $\omega_3 = 0.1$ .

**4.2. Result Analysis.** In this paper, according to the parameters set in Section 4.1, the deep neural network training is carried out. The trained network model is used in the MEC system simulation experiment, and the simulation time is set to 300 minutes. In order to better reflect the effect of algorithm optimization, under the same experimental conditions, the proposed algorithm is compared with Ref. [8] algorithm and Ref. [9] algorithm. In this paper, user request receipt rate, total number of migrated sessions, and running time are used as performance evaluation indicators.

Set the video content requested by the user to follow Zipf distribution. Figures 5–7 show the relationship between the user request reception rate, the total number of migration sessions and the running time of the simulation algorithm, and the system load under this condition, respectively.

As can be seen from Figures 5 and 6, in the case of low load ( $\zeta \leq 61$ ), the user request reception rate and the total number of migration sessions of this method are basically the same as Ref. [8] algorithm and Ref. [9] algorithm. In the case of high load ( $\zeta > 61$ ), the receiving rate of user requests and the total number of migrating sessions of this method are lower than Ref. [8] algorithm and Ref. [9] algorithm.

Compared with Ref. [8] algorithm and Ref. [9] algorithm, the average receipt rates of user requests in this method are reduced by 0.85% and 1.72%, respectively, and the total number of migrated sessions is reduced by 3.55% and 5.29%, respectively. The result shows the advantage of reinforcement learning. Because session transfer is cost-effective, in order to obtain greater returns, the decision-maker constantly adjusts the decision-making actions and ultimately reduces the cost of transfer, while guaranteeing a higher request reception rate.

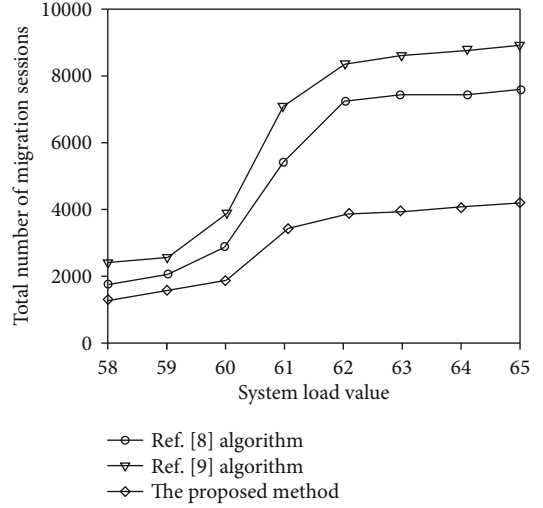


FIGURE 10: The relation between running time and system load under random uniform distribution.

As can be seen from Figure 7, for both low load and high load, the running time of the proposed algorithm is better than Ref. [8] algorithm and Ref. [9] algorithm, and the running time is shortened by 39.98% and 54.54% on average, respectively. Because in the process of user requesting access, the session allocation threshold needs to be constantly updated by method Ref. [8] algorithm, which leads to a lot of computation, and method Ref. [9] algorithm needs to be constantly overlapped. In order to find the optimal solution, the deep reinforcement learning method used in this paper only needs to make scheduling decisions through the trained strategy network, which has less computational complexity and improves efficiency.

In order to evaluate the adaptability of the proposed algorithm, a random uniform distribution of video content requested by users is set up. Figures 8–10 show the relationship between the request reception rate, the total number of migration sessions, and the running time of the simulation algorithm and the system load under this condition, respectively. Compared with Ref. [8] algorithm and Ref. [9] algorithm, the average receipt rate of user requests in this algorithm is reduced by 0.41% and 1.19%, the total number of migrated sessions is reduced by 3.64% and 6.57%, respectively, and the running time is reduced by 45.28% and 56.03%, respectively. The experimental results show that the proposed algorithm has a certain degree of self-adaptability. When the distribution of user requests changes, the scheduling strategy can still be adjusted in the training process, resulting in a lower migration cost and a higher user request reception rate.

In summary, the proposed deep reinforcement learning-based session scheduling strategy for streaming media edge cloud not only achieves better request access effect but also has lower migration cost. More importantly, it has a great speed advantage, that is, shorter running time. At the same time, it has strong adaptability in an uncertain MEC system environment.

## 5. Conclusion

In order to achieve efficient and smooth resource scheduling for streaming media service system in cloud mode, this paper proposes a video stream session migration method based on deep reinforcement learning. The method transforms session migration problem into reinforcement learning problem; defines state space, action set, and return function; calculates session volume according to load; and uses convolutional neural network to fit behavior selection strategy function and action-value function. The experimental results show that compared with the methods of Ref. [8] algorithm and Ref. [9] algorithm, this strategy can reduce the migration cost and shorten the running time.

This paper only considers the video session request access server as the output of  $Q$  network. Later research focuses on the video session request access server and the video session moved out of the server as the output of  $Q$  network, in order to improve the migration method of streaming media edge cloud session and extend the application object to dynamic video session.

## Data Availability

The data included in this paper are available without any restriction.

## Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

## Acknowledgments

We wish to express their appreciation to the reviewers for their helpful suggestions which greatly improved the presentation of this paper. This work was supported by the Henan Province Science and Technology Project (142102210366), Henan Public Security Think Tank Project (No. 2020-25), and Henan Police College Education and Teaching Reform Research and Practice Project (No. 2020-9).

## References

- [1] K. Kaur, S. Garg, G. S. Aujla, N. Kumar, J. J. P. C. Rodrigues, and M. Guizani, "Edge computing in the industrial internet of things environment: software-defined-networks-based edge-cloud interplay," *IEEE Communications Magazine*, vol. 56, no. 2, pp. 44–51, 2018.
- [2] R. Kiruthiga and D. Akila, "Heterogeneous fair resource allocation and scheduling for big data streams in cloud environments," in *2021 2nd International Conference on Computation, Automation and Knowledge Management (ICCAKM)*, pp. 128–132, Dubai, United Arab Emirates, 2021.
- [3] P. Pradhan, P. K. Behera, and B. N. B. Ray, "Improved max-min algorithm for resource allocation in cloud computing," in *2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC)*, pp. 22–24, Wagnaghat, India, 2020.
- [4] E. Korpeoglu, C. Sahin, D. Agrawal, A. E. Abbadi, T. Hosomi, and Y. Seo, "Dragonfly: cloud assisted peer-to-peer architecture for multipoint media streaming applications," in *2013 IEEE Sixth International Conference on Cloud Computing*, pp. 269–276, Santa Clara, CA, USA, 2013.
- [5] J. Tongquan, Z. Wang, and X. Hongsheng, "Session migration strategy for streaming media edge cloud based on dynamic threshold allocation," *Computer Engineering*, vol. 43, no. 1, pp. 55–60, 2017.
- [6] I. Baumgart, B. Heep, and S. Krause, "Over Sim: a scalable and flexible overlay framework for simulation and real network applications," in *2009 IEEE Ninth International Conference on Peer-to-Peer Computing*, pp. 87–88, Seattle, WA, USA, 2009.
- [7] R. S. Sutton, D. Precup, and S. Singh, "Between MDPs and semi-MDPs: a framework for temporal abstraction in reinforcement learning," *Artificial Intelligence*, vol. 112, no. 1–2, pp. 181–211, 1999.
- [8] T. Jiang, Z. Wang, Z. Chen, and H. Xi, "An adaptive strategy of online session migration for Streaming Media Edge Cloud," in *2016 35th Chinese Control Conference (CCC)*, pp. 5278–5283, Chengdu, China, 2016.
- [9] Z. Chen, Z. Wang, and H. Xi, "A dynamic two-phase schedule strategy for streaming media edge cloud," in *2017 29th Chinese control and decision conference (CCDC)*, pp. 2281–2286, Chongqing, 2017.
- [10] R. Viola, A. Martin, M. Zorrilla, and J. Montalbán, "MEC proxy for efficient cache and reliable multi-CDN video distribution," in *2018 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, pp. 1–7, Valencia, Spain, 2018.
- [11] M. Liu, Y. Teng, F. R. Yu, V. C. M. Leung, and M. Song, "A mobile edge computing (MEC)-enabled transcoding framework for Blockchain-based video streaming," *IEEE Wireless Communications*, vol. 27, no. 2, pp. 81–87, 2020.
- [12] Y. Liu, B. Du, S. Wang, H. Yang, and X. Wang, "Design and implementation of performance testing utility for RTSP streaming media server," in *2010 First International Conference on Pervasive Computing, Signal Processing and Applications*, pp. 193–196, Harbin, China, 2010.
- [13] H. Wang, J. Li, C. Zhao, and Z. Ying, "Design of an embedded streaming media server in video monitoring," in *2013 Ninth International Conference on Natural Computation (ICNC)*, pp. 1324–1328, Shenyang, China, 2013.
- [14] Q. Fan and X. Wang, "Design of streaming media server based on softswitch platform," in *2012 IEEE International Conference on Computer Science and Automation Engineering (CSAE)*, pp. 288–290, Zhangjiajie, China, 2012.
- [15] A. Argyriou, "Link scheduling for multiple multicast sessions in distributed wireless networks," *IEEE Wireless Communications Letters*, vol. 2, no. 3, pp. 343–346, 2013.
- [16] Yongfei Zhang, Shiyin Qin, and Zhihai He, "Fine-granularity transmission distortion modeling for video packet scheduling over mesh networks," *IEEE Transactions on Multimedia*, vol. 12, no. 1, pp. 1–12, 2010.
- [17] S. Zhang and C. Chan, "Provisioning of survivable multicast sessions in wavelength-routed optical networks with scheduled traffic," *Journal of Lightwave Technology*, vol. 29, no. 5, pp. 685–690, 2011.
- [18] Y. Han, Z. Wang, S. Chen, G. Li, X. Zhang, and X. Yuan, "Interactive assigning of conference sessions with visualization and topic modeling," in *2020 IEEE Pacific Visualization Symposium (Pacific vis)*, pp. 236–240, Tianjin, China, 2020.

- [19] W. Kuo and C. Wang, "Robust and optimal opportunistic scheduling for downlink 2-flow inter-session network coding with varying channel quality," in *IEEE INFOCOM 2014- IEEE Conference on Computer Communications*, pp. 655–663, Toronto, ON, Canada, 2014.
- [20] B. Kim and J. Lee, "Opportunistic resource scheduling for OFDMA networks with network coding at relay stations," *IEEE Transactions on Wireless Communications*, vol. 11, no. 1, pp. 210–221, 2012.

## Research Article

# Multiobjective Optimization Method of Coevolution to Intelligent Agricultural Dynamic Services under the Internet of Things Environment

Haihong Liang 

*The Department of Business Administration, Zhejiang Economic & Trade Polytechnic, Hangzhou, Zhejiang 310033, China*

Correspondence should be addressed to Haihong Liang; [lh@zjiet.edu.cn](mailto:lh@zjiet.edu.cn)

Received 27 January 2021; Revised 9 March 2021; Accepted 22 March 2021; Published 12 April 2021

Academic Editor: Shan Zhong

Copyright © 2021 Haihong Liang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The agricultural Internet of Things system, with its large-scale, highly heterogeneous, and dynamic characteristics, brings certain difficulties to the provision of agricultural Internet of Things services. Considering the multiple requests of the agricultural Internet of Things at any random moment, which have the characteristics of multiple sources, multiple types, and uneven tasks, this paper establishes an optimization model for the minimum service cost and proposes a collaborative evolution to intelligent agricultural dynamic services under the Internet of Things environment multiobjective optimization method. First, according to the probability that the allele on the fragment to be vaccinated has appeared in the memory bank, use the detection strategy to judge whether the solution is illegal; secondly, compare the optimal individual with other values appearing on the gene locus, judge whether the optimal gene or fall into the local optimal, and inoculate with probability through simulated annealing; finally, the total service cost and service time were evaluated under the two service provision strategies and compared with the other three intelligent algorithms; the results confirmed the feasibility and effectiveness of the proposed algorithm. At the same time, the simulation results show that the proposed collaborative multiobjective optimization algorithm can achieve better performance.

## 1. Introduction

The Internet of Things is a technological form with the characteristics of the times. It is a comprehensive technological form that will inevitably arise after network technology, embedded technology, sensor technology, cloud computing technology, and other related technologies that reach a certain level. It is an essential data acquisition, transmission, and processing mode in the era of big data [1]. As a general technical means, the Internet of Things has been extensively studied. However, the combination of the Internet of Things and specific fields requires reoptimization and design of the technology [2–4].

The agricultural Internet of Things is the core technology of the smart agricultural system. The smart agricultural system is a complex system formed by the fusion of multiple systems [5]. Errors in system design can have disastrous consequences for production. In the process of applying the

Internet of Things to the agricultural system, the correctness of the system design directly affects the operation of the system [6, 7]. Regardless of hardware or software, due to the complexity of the system itself and the diversity of the agricultural environment in which it is located, it will be difficult to achieve the perfect system design. How to ensure the correctness of the design of the agricultural Internet of Things system is a problem that must be solved [8]. The agricultural Internet of Things connects the three parties of “man-machine-things” to form a community. It is helpful for agricultural practitioners to manage and control all links, elements, and subsystems in agriculture more scientifically, precisely, and in real time. This greatly improves human’s awareness of the individual nature of agricultural animals and plants, the monitoring and adjustment capabilities of complex agricultural systems, and the ability to respond to emergencies in agricultural production [9]. The agricultural Internet of Things is a systematic project with strong



comprehensiveness. It is the only way for agricultural production to achieve green environmental protection, high-quality, high-efficiency, and scientific development. The traditional agricultural model is far from being able to meet the long-term sustainable development needs of agriculture. Currently, research on the Internet of Things in agriculture has been widely implemented worldwide. But overall, the application is still in the experimental demonstration stage [10–12].

The agricultural Internet of Things service is an emerging industry and an inevitable trend of intelligent agricultural production in the future [13]. From the perception of the Internet of Things to decision-making and then to control, it is a necessary link to realize the integrated intelligent control of the autonomous service of the Internet of Things. The perception of the Internet of Things has problems such as information uncertainty and high redundancy. It is necessary to study data processing methods such as data cleaning, data compression, and data fusion to realize the efficient perception of Internet of Things information [14–16]. Sensor perception is like the human sensory system, actively discovering or passively receiving external stimuli. They can extract details from the external environment even faster than human observers.

Since artificial intelligence was proposed in 1956, it has been the research direction of scholars and computational intelligence is an important branch of artificial intelligence research. Computational intelligence is based on the perspective of biological evolution to recognize and simulate intelligence. It is based on data and solves problems through training to establish connections [17–19]. Natural biological systems have evolved many mechanisms in the process of continuous evolution and have strong adaptability to complex and dynamic changes in the external environment. The different components or organizations of the biological system can function autonomously and collaboratively under limited rules to realize the complex behaviors and functions of the entire system. At present, many scholars try to summarize and abstract the action rules and regulatory control mechanisms of biological systems into mathematical forms or action rules. It provides a theoretical basis and a new source of inspiration for designing efficient and robust complex network architectures, control systems, learning systems, and optimization algorithms [20].

## 2. Related Works

The application scenarios of the agricultural Internet of Things are mainly concentrated in the field, facility gardening, and livestock and poultry breeding. Since the agricultural Internet of Things is based on a specific application scenario of agriculture, it should be closely integrated with agriculture-related scenarios in its design and implementation. In the realization of the system, the universal characteristics of the Internet of Things should be considered and the characteristics of the combination with agricultural production conditions should be considered. The data collected by these perception layers are transmitted through the Internet of Things. In the application management layer, the data is

analyzed and decided. Participate in the automatic control and adjustment of the production environment of agricultural products to ensure that agricultural products have a suitable production environment and obtain the best production conditions.

Reference [21–23] discussed the issue of resource service provision on cloud platforms. Efficient resource service provision strategy can guarantee satisfactory cloud computing services to end users. Reference [24] introduced the Nash equilibrium to deal with resource transaction activities between cloud service providers and used game theory in the process of providing Internet of Things services. Reference [25] uses resource sharing among cloud service providers for customs clearance to achieve efficient use of limited resources and improve service quality. In the cloud model, on-demand computing resources (such as networks, storage, and servers) can be allocated from a shared resource pool with minimal management. Infrastructure as a service (IaaS) is one of the most common cloud service models. In the IaaS model, services are deployed as virtual machines in the cloud computing infrastructure, with the goal of extreme use of resources [26]. In contrast, in the resource and service scheduling model widely studied in cloud computing, users are charged according to their resource usage and required service quality specifications, while the focus of the Internet of Things service model is different from the cloud model. Although cloud computing is important in Internet of Things analysis and data storage, the Internet of Things does not necessarily include the cloud. Nevertheless, the service scenarios are different but the ideas provided by these services can open up ideas for the provision of Internet of Things services.

Reference [27] proposed a service provision platform for the Internet of Things based on a multilevel and multidimensional model. It can access large heterogeneous resources and expose their functions as a lightweight service. In a distributed Internet of Things environment, it provides a unified message space to promote the on-demand dissemination and sharing of sensor information. The platform supports application sharing and reuse of resources and provides an infrastructure for Internet of Things application models. Reference [28] uses the neuroendocrine system's comprehensive regulation mechanism of blood glucose concentration to realize the monitoring of randomly generated service requests by the Internet of Things. However, the service delivery strategy has not been studied in depth. Reference [29] proposed a framework model for service submission based on the Internet of Things. Based on this model, abstract and dynamic collaboration methods are used to realize the resource sharing of the Internet of Things and the service dynamic composition technology is adopted to realize the rapid submission of complex services. However, it does not consider the algorithm's resource consumption and optimization of the service capability itself. Reference [30] adopts a method of network perception, using genetic algorithm to realize adaptive service selection and provision. Reference [31] adopts configurable event-driven combination and adaptive service provision, which is divided into three core services: coordination service, context service, and event service.

At this stage, the agricultural Internet of Things is in the early stage of construction, commercial terminals have not yet been widely popularized, the business volume is small, and there is a lack of optimization experience to learn from. It is urgent to carry out targeted research on network and business collaborative optimization. The main research contents of this paper are as follows:

- (1) Based on the probability that the alleles on the fragments to be vaccinated have appeared in the memory bank, the detection strategy is used to judge whether the solution is illegal. It can effectively improve the shortcomings of premature convergence and poor local search ability
- (2) Compare the optimal individual with other values that have appeared on the gene locus, and judge whether the optimal gene or the local optimal. And inoculate with probability through simulated annealing to improve the robustness of the algorithm
- (3) Evaluate the total service cost and total service time under the two service provision strategies. The experimental results verify the feasibility and effectiveness of the proposed algorithm

### 3. Agricultural Internet of Things Service Architecture and Optimization Model

*3.1. Agricultural Internet of Things Service Architecture.* There are two main occasions for agriculture: facility agriculture and field. In facility agriculture, there are usually more greenhouse environments. The research in this paper is aimed at coevolutionary multiobjective optimizing the methods for intelligent agricultural dynamic services in the Internet of Things environment. Various sensors are installed to monitor the growth of various vegetables and fruits. The service architecture of the agricultural Internet of Things is shown in Figure 1.

Examples of sensors include temperature sensors, humidity sensors, soil moisture sensors, nutrition measurement sensors, and carbon dioxide sensors, used to measure physical parameters such as temperature, relative humidity, soil water content, soil nutrients, and carbon dioxide concentration. These monitored data are sent to the computer platform (network layer) through aggregation and then analyzed and processed. Data that meets a certain condition is considered a service request in this chapter. When the sink node integrates all kinds of data from sensors, several service requests will be generated. All the request tasks are scheduled and assigned to different service devices by the service computing platform in a global optimization way. Here, we consider them as a service request queue. The multiservice window system of agricultural Internet of Things can be modeled as the  $M/G/n$  queuing model. This paper adopts the parallel batch processing mode for requests in unit time. In this agricultural Internet of Things service scenario, some agricultural equipment, such as irrigation machines, fertilizer spreaders, and spraying machines, is a service provider. The agricultural Internet of Things service architecture has three

layers: intelligent perception layer, network layer, and equipment control layer. Suppose that there are only two types of equipment that provide services, irrigation machines and fertilizer spreaders. According to the degree of water shortage or undernutrition, the priority of requests is defined into two categories: urgent and general. If this is an urgent task, it will have a higher priority.

The most basic components of an Internet of Things temperature sensing service should be completed by related functional components such as sensors, Internet of Things data transmission layers, and web services in the Internet of Things system. At the data transmission layer of the Internet of Things, it is possible to complete the encapsulation of the Internet of Things functions required for temperature sensing services, such as combining the relative position information of the humidity sensor node with the temperature value to form a new data structure and other operations. The release of temperature sensing service can be done by using the web service in the system, and the web service can provide functions such as retrieval and browsing of related information.

*3.2. Optimization Model.* The biological immune system is a kind of control system with strong robustness and adaptive ability in complex disturbance and uncertain environment. The endocrine system and the immune system have a close two-way regulatory connection. Endocrine hormones can enhance or weaken the immune function through hormone receptors on immune cells. The immune system affects the endocrine system through cytokines, and the same cytokine has specificity in the regulation of different hormone secretion activities.

There is a general rule of hormone secretion by hormone glands. The rise and fall of hormone regulation follows the Hill function, as shown in formulas (1) and (2), respectively.

$$F_{\text{up}}(G) = \frac{G^n}{G^n + T^n}, \quad (1)$$

$$F_{\text{down}}(G) = \frac{T^n}{G^n + T^n}, \quad (2)$$

where  $G$  is the independent variable,  $T$  is a threshold value greater than 0,  $n$  is the Hill coefficient, and its value is greater than 1.  $n$  and  $T$  jointly control the slope.

Using the Hill function, the secretion of hormone  $A$  is expressed in the following form:

$$S_A = aF_{\text{up(down)}}(B) + S_{A0}, \quad (3)$$

where hormone  $B$  controls hormone  $A$ ,  $S_{A0}$  is the base value of hormone  $A$ , and  $a$  represents a constant coefficient.

AIE is an improved adaptive immune algorithm based on the endocrine regulation mechanism. This algorithm can make individuals have good diversity and effectively overcome the problems of immature convergence and slow evolution. This is mainly due to the adaptive crossover and mutation probability factors designed in this paper. Follow the law of Hill function to regulate the rise and fall of endocrine hormone secretion regulation.

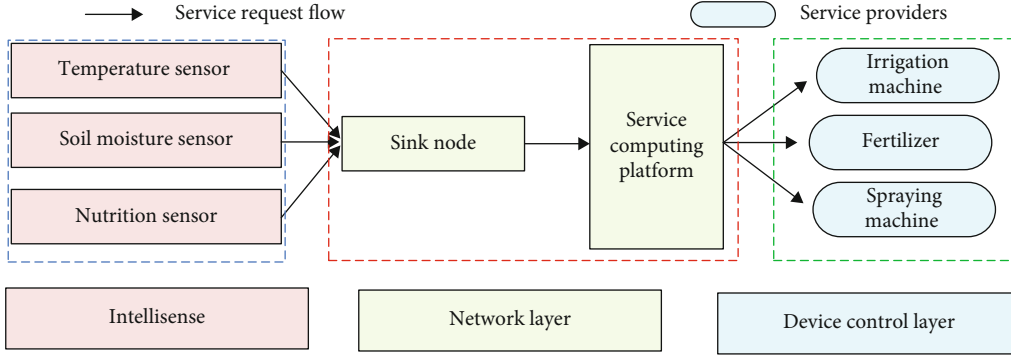


FIGURE 1: Agricultural Internet of Things service architecture.

#### 4. Adaptive Immune Algorithm Based on Endocrine Regulation

4.1. *The Mapping Relationship between the Agricultural Internet of Things System and the Immune System.* Immunity is a part of species evolution, which enables the body to distinguish between the antigen and the body in the body and produce antibodies to inhibit the invasion of the antigen. After the action is over, the antibodies will be balanced to keep the environment in the body stable. Incidentally, immune cells are produced to produce rapid and effective responses to similar antigens in the future. The process by which organisms produce antibodies to foreign antigens through biochemical action to eliminate the invasion is called immune response. During the first immune response, the information is retained by memory cells after the body clears the antigen. When the antibody carrying this kind of information invades again, it will activate the memory cells and quickly produce equivalent antibodies to eliminate the effect through cell differentiation. This phenomenon has reference significance for optimization problems.

4.2. *Adaptive Immune Algorithm Flow.* The adaptive thought of genetic algorithm refers to an algorithm that dynamically adjusts the control parameters in the algorithm according to the changes in the fitness of the population during the operation of the genetic algorithm. Its main purpose is to increase the diversity of the population at the right time to avoid the situation where the algorithm falls into a local optimal solution. As far as genetic algorithm is concerned, it is mainly used to adjust the crossover probability and mutation probability.

This paper mainly uses information entropy to characterize the similarity of the population, so as to adaptively adjust the crossover probability and mutation probability. The specific description of the algorithm is as follows:

- (1) Population average information entropy

$$H(n) = \frac{1}{L} \sum_{j=0}^{L-1} H_j(N), \quad (4)$$

where  $H_j(N)$  is the information entropy of the  $j$ -th gene:

$$H_j(N) = \sum_{n=0}^{S-1} -p_{ij} \log_2 p_{ij}, \quad (5)$$

where  $p_{ij}$  is the probability that the  $j$ -th symbol appears on the locus

- (2) Population similarity

$$A_{(N)} = \frac{1}{1 + H(N)}, \quad (6)$$

where  $A_{(N)}$  characterizes the overall similarity of the entire population

- (3) Adaptive strategy

When the fitness of the population tends to be consistent or tends to the local optimum, increase  $P_c$  and  $P_m$ , and decrease  $P_c$  and  $P_m$  when the fitness is relatively scattered.  $P_c$  and  $P_m$  are automatically adjusted according to the following formula:

$$\begin{aligned} P_c &= e^{2(A(N)-1)}, \\ P_m &= 0.1e^{2(A(N)-1)} \end{aligned} \quad (7)$$

- (4) Adaptive extraction of vaccines

The adaptive extraction of vaccines specifically includes two processes: extraction of vaccines and vaccination. Assuming that the  $K_b$  antibodies with the best fitness in each generation preserved population, then  $K * K_b$  antibodies were retained in the most recent  $K$  generation to form the optimal antibody group. Each locus of each antibody has  $k_1, k_2, \dots, k_s$ , where  $s$  symbols for selection. The probability that the  $i$ -th allele is  $k_j$  is

$$p_{ij} = \frac{1}{K * K_b} \sum_{j=1}^{K * K_b} a_j \quad (8)$$

Select the antibody to be vaccinated from the parent population, and select one or more gene fragments according to the roulette method. Generate new immune individuals by replacing gene code values to form a better population. The gene fragment selection method is as follows: from the above extraction of the vaccine, we know that the vaccine is  $H = (h_1, h_2, \dots, h_L)$ , so

$$q_i = \frac{p_i}{\sum_{j=1}^L p_j}, \quad i = 1, 2, \dots, L. \quad (9)$$

This paper proposes an adaptive immune genetic algorithm with endocrine regulation (AIGE). The information entropy is calculated to evaluate the similarity of the population, so as to adaptively adjust the crossover probability and mutation probability. The algorithm adopts the method of dynamically extracting vaccines, which avoids the shortcoming of a slower convergence rate due to the reduced effectiveness of static vaccines in the evolution process. The flow of the improved adaptive immune genetic algorithm is shown in Figure 2.

**4.3. Improved Design of the Immune Genetic Algorithm.** The simple immune genetic algorithm simulates the antibody memory antigen information bionic from the immune system of the organism into a vaccine operator composed of recording tape solving problem feature information, prior knowledge, and heuristic rules. By extracting vaccines, the original random search of the genetic algorithm is improved to some extent. This improvement has demonstrated excellent search results in many combination optimization problems. But with the continuous optimization of intelligent algorithms, immune genetic algorithm, as an improved algorithm of genetic algorithm, still inherits randomness. The cross-based global search weakens the local convergence, and the Pareto solution is often not obtained when solving large-scale problems with complex space. Because vaccination and immune selection tend to converge in one direction, it is not conducive to maintaining diversity of the population.

The survival of the fittest in the evolution process is judged based on the affinity set by the objective function. It is impossible to absolutely judge the pros and cons of individual gene segments, leading to missed opportunities to find the optimal solution earlier. The setting of parameters needs to be considered for various algorithms. Setting fixed control parameters for operations such as crossover and mutation before searching is not suitable for the updated status of the new population. The extraction of vaccines is also difficult. Insufficient and inaccurate prior knowledge will reduce the overall effectiveness of the algorithm. The implicit information of a certain vaccination may lead to iterative guidance, causing the search to go in a certain direction and stall. In addition, the preextracted vaccine sometimes cannot obtain prior knowledge from the question to be asked or it is difficult to discover characteristic information. Taking too long to

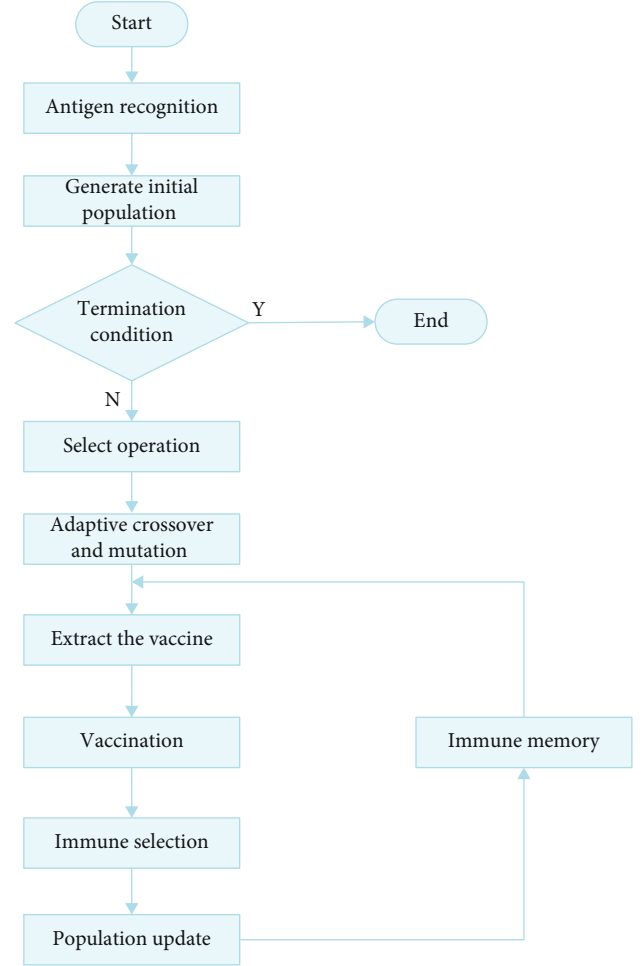


FIGURE 2: Flow chart of the adaptive immune genetic algorithm. Y: yes; N: no.

extract information will lead to the time inefficiency of the algorithm. Aiming at these deficiencies of the basic immune genetic algorithm, this paper proposes to improve the algorithm design in the following aspects: adopting a two-layer coding based on machine selection and process sequencing to avoid illegal solutions.

Evaluation the antibodies after population initialization. Using concentration inhibition, the concentration  $C_v$  of antibody  $v$  in the population refers to the proportion of the antibody with its similar antibodies in the population, namely,

$$C_v = \frac{1}{N} \sum_{w=1}^N A_{vw}. \quad (10)$$

In population renewal, the concentration of antibodies with high affinity is increased and it will be inhibited when it reaches a certain value. On the contrary, the production and selection probability of antibodies with low concentration should be increased. Therefore, when judging whether the diversity meets the requirements, different similarity thresholds  $\gamma$  should be set according to the size of the group  $N$ . But so far, the selection of  $\gamma$  still lacks a strong theoretical

basis. Refer to the fitness threshold of the critical path to modify it so that it quickly converges and obtains the global optimal solution.

Refer to medical treatment to prevent allergic reaction, immune deficiency, autoimmunity, virus infection and other immune abnormalities. For vaccination, the “skin test” and other prevaccination tests are carried out to avoid overimmunity and underimmunity. The allele selection of the detection strategy is used to judge. The specific method is to compare the corresponding loci of the new and old optimal individuals. If the probability of occurrence is small, strictly control it and decide whether to discard it after judging whether the solution is illegal. If the similarity of a certain gene locus to be vaccinated is high, it is also judged whether it is the best gene or falls into the local optimum. If the gene locus has other values that can constitute an individual with higher fitness, simulated annealing is used to inoculate with probability to avoid local optimality. The method is as follows: when the temperature is  $T$ , the  $i$ -th allele  $K_i$  to be vaccinated is  $h_i$  and the difference between the two solutions is  $\Delta f = f(h_i) - f(K_i)$ . If  $\Delta f < 0$ , inoculate; if  $\Delta f > 0$ , inoculate with probability  $P_v$

$$P_v = T(t) * e^{\frac{f(h_i) - f(K_i)}{kT}}, \quad (11)$$

where  $T(t) = k * T(t-1)$ ,  $t$  is the cooling times, and Boltzmann constant  $k$  can be set as a positive number close to 1, such as 0.95.

**4.4. Convergence Analysis.** Let the population size of the immune genetic algorithm be  $N_0$ . Let the length of individual chromosomes be  $L$ . Genes are coded in the  $Q$  system, and the crossover and mutation probabilities are  $P_c$  and  $P_m$ , respectively. The probability of a certain gene mutation in another situation is  $1/(Q-1)$ . The state transition equation of immune genetic algorithm is

$$A_k \xrightarrow{\text{cross}} B_k \xrightarrow{\text{mutations}} C_k \xrightarrow{\text{vaccination}} D_k \xrightarrow{\text{immune selection}} A_{k+1}. \quad (12)$$

The state transition from  $A_k$  to  $D_k$  forms a Markov chain. By analogy,  $A_{k+1}$  to  $D_{k+1}$  are still a Markov chain. Let  $N_0$  be a point in the state space  $S = X^{N_0}$ ; then, the coordinates of the point are individuals in the search space  $X$ . The number of  $S$  states is denoted by  $|S|$ . A certain state in  $S$  is denoted as  $S_i \in S$ ,  $i = 1, 2, \dots, |S|$ . The ownership of the subset in the search space is represented by  $S_i \in S_j$ . Use  $V_k^i$  to represent the state  $s_i$  of the random variable  $V$  in the  $k$  generation.

Let  $f$  be the fitness function on  $X$ . Suppose that the optimal state set  $S^* = \{x \in X | f(x) = \max_{x_i \in X} f(x_i)\}$ . For any population distribution after multiple iterations, the probability convergence is expressed as

$$\lim_{k \rightarrow \infty} \sum_{s_i \cap S^* \neq \emptyset} P\{A_{k+1}^i\} = 1. \quad (13)$$

Research has proved that the immune genetic algorithm

TABLE 1: Regional settings of agricultural Internet of Things.

Parameter	Value
Length	90
Width	90
Number of device nodes	100
Number of sensor nodes	200

TABLE 2: Poisson distribution parameter setting.

Parameter	Value
$\lambda = 5$	$R = [3, 5, 8]$
$\lambda = 10$	$R = [8 - 11, 13]$
$\lambda = 15$	$R = [18, 10, 15, 32, 8]$
$\lambda = 20$	$R = [28, 31, 26, 35, 32]$

TABLE 3: Each algorithm parameter setting.

Parameter	PSO	GA	IA	Improved AIE
Population size	60	60	60	60
Crossover probability	/	0.25	0.6	Adaptive
Mutation probability	/	0.02	0.1	Adaptive
$p_c^0$	/	/	/	0.6
$p_m^0$	/	/	/	0.2
$\alpha$	/	/	/	0.1
$\beta$	/	/	/	0.2
$n_c$	/	/	/	2
$n_w$	/	/	/	2
$\omega$	[0.3, 0.88]			/
$c_1, c_2$	[0.8, 2.7]	/	/	/

is probabilistically convergent, and the genetic algorithm alone cannot guarantee convergence to the global optimal solution. Vaccines that improve antibody fitness above the average level will spread exponentially and vice versa.

## 5. Experiment and Result Analysis

**5.1. Parameter Settings.** This section designates an agricultural Internet of Things area for experimentation. The perception data from the sensor needs to be analyzed and processed according to the various parameters of good plant growth. After the service request is issued, the service computing platform calculates and selects the optimal service device based on the request type and workload and then turns on and controls the device switch.

The parameter value setting is shown in Table 1. In Table 2,  $\lambda$  indicates the average number of requests that reach the system per unit time. In this section, five batches of the request tasks are tested under each  $\lambda$  value. The number of requests in each batch is based on random numbers generated under the current  $\lambda$  value. The number of sensor nodes

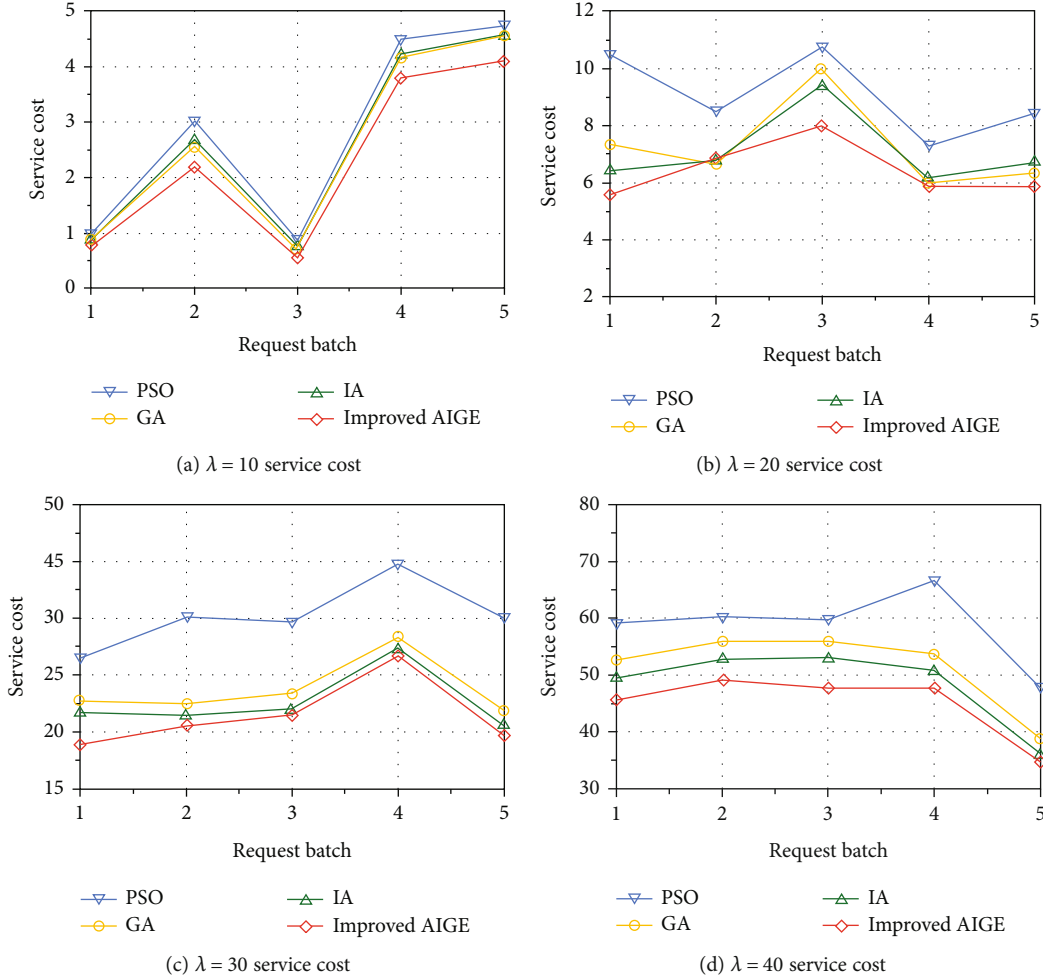


FIGURE 3: Comparison of objective function values under a single-service strategy.

is set to 200 to ensure a sufficient number of requests in different experimental environments. The number of device nodes is set to 100. Considering the actual scenario, the greenhouse generally uses nearby irrigation machines or other equipment. If the distance between the request and the service is greater than a certain set threshold, the obtained service plan is meaningless. Therefore, the optimal service scheduling schemes given in this article are all based on the scope of implementation of the equipment.

Table 3 gives the parameter values when various algorithms are running. The termination conditions of each algorithm are set to 100 epochs.

**5.2. Comparison of Simulation Results.** In order to fully verify the performance advantages of the proposed algorithm for improving AIGE, this paper chooses particle swarm optimization (PSO), genetic algorithm (GA), immune algorithm (IA), and the algorithm of this paper to compare experiments.

Figure 3 shows the total service cost of different request arrival rates under the single-service strategy. In Figures 3(a)–3(d), it can be observed that the adaptive immune algorithm proposed in this paper has the best performance under four different request arrival rates. When  $\lambda$

$= 5$  and  $\lambda = 10$ , the proposed improved AIE algorithm is better than the PSO algorithm, which is roughly consistent with the performance of the GA algorithm. The results also show that as the value of  $\lambda$  increases, the proposed improved AIE algorithm gradually shows its superiority over GA and PSO algorithms. This shows that the proposed improved AIGE algorithm has the potential to solve high-dimensional problems.

Figure 4 shows the total service cost of different request arrival rates under the collaborative service strategy. As shown in the figure, the algorithm proposed in this paper has the lowest service cost ratio and its performance ranks first. It can be seen in Figure 4(a) that when  $\lambda = 10$ , the total cost value obtained by improving the AIE algorithm and GA is very close. It can be seen in Figure 4(d) that the difference between the improved AIGE algorithm and the results obtained by GA in processing the fourth and fifth batches of requests is relatively small. This is because their ability to explore the solution space is close when requesting these two large quantities under the collaborative service strategy. When the number of requests in the first three batches is relatively small, the improved AIE algorithm is slightly better than the GA algorithm. In Figures 4(b) and 4(c), it can be observed that the improved AIGE algorithm

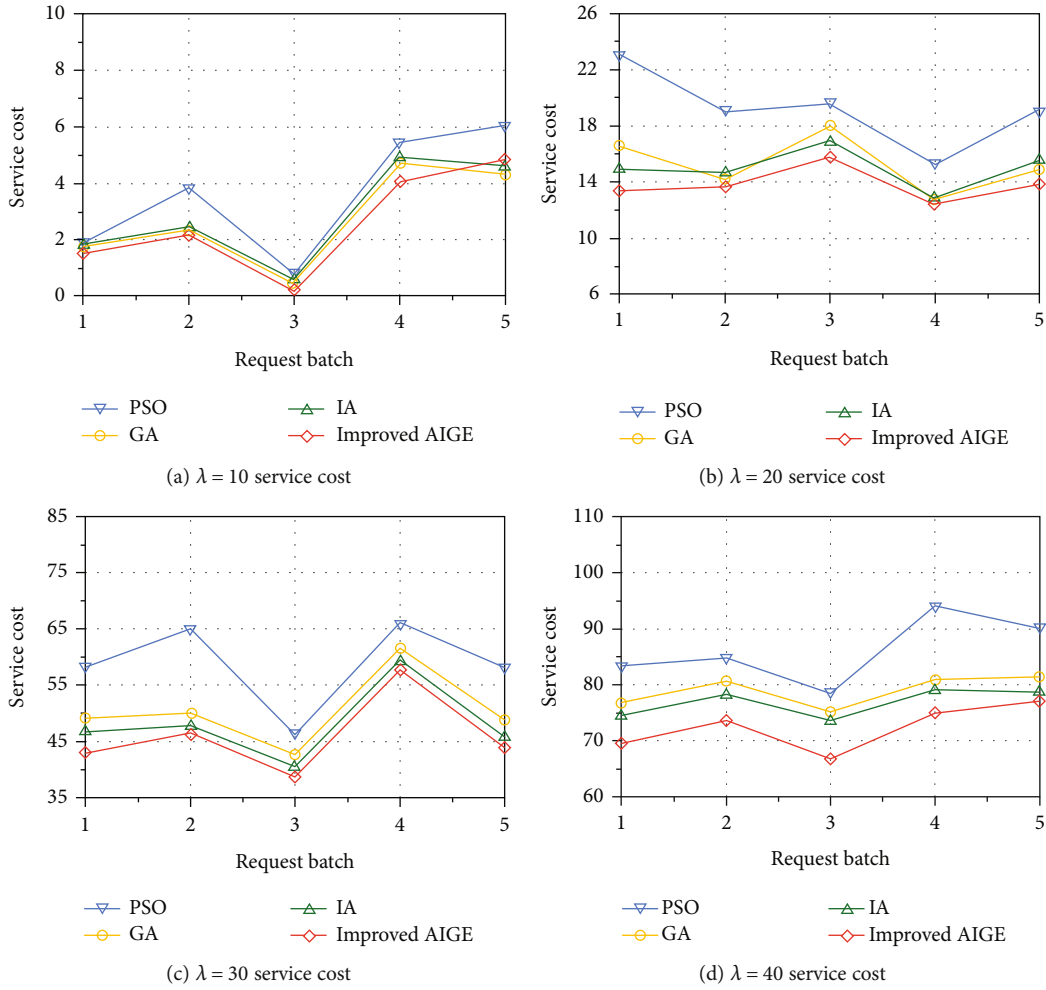


FIGURE 4: Comparison of objective function values under the collaborative service strategy.

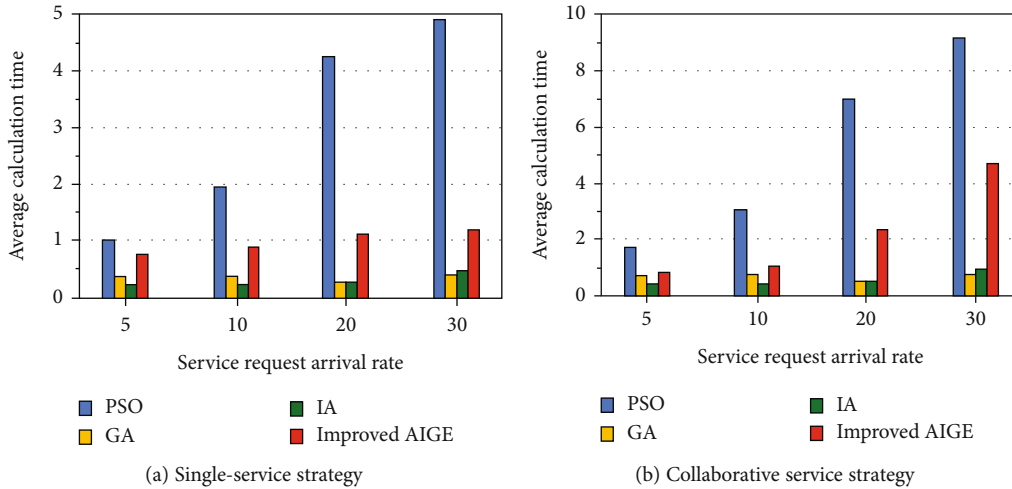


FIGURE 5: Average calculation time under different request arrival rates.

has the best performance. When  $\lambda=30$  and  $\lambda=40$ , the results shown in Figures 4(c) and 4(d) reveal the superiority of the improved AIGE algorithm under the collaborative service strategy.

It can be seen from the above results that the total cost curve conforms to normal fluctuations, which depends on the change in the number of requests. For the same  $\lambda$  value, the change directions of the curves of the four algorithms

are the same. The service system can choose the service mode freely according to the specific situation. For example, when multiple requests are in the peak period or the task is urgent, the system can choose the cooperative service strategy to reduce the service time; otherwise, it can choose the single-service strategy.

It can be seen in Figure 5(a) that for four different  $\lambda$  values, the curves of the GA algorithm and the IA algorithm remain relatively smooth. As the value of  $\lambda$  increases, the curve of the improved AIGE algorithm shows a slow upward trend, while the PSO shows a sharp upward trend. This illustrates that for PSO, the calculation time is very sensitive to the dimensionality of the problem. It can be seen in Figure 5(b) that each algorithm under the collaborative service strategy takes more time than under the single-service strategy. Because the collaboration strategy requires more service providers to participate in the task. The increase in computing workload is normal. Although the improved AIGE algorithm proposed in this paper consumes more computing time than the IA algorithm and the GA algorithm, it obtains a lower service cost.

## 6. Conclusion

This article systematically introduces and analyzes the service selection and optimization methods of the agricultural Internet of Things. Collaborative optimization of Agricultural Internet of things service resources has been deeply studied. A single-service strategy and a collaborative service strategy optimization model are constructed. A coevolutionary multi-objective optimization method for intelligent agricultural dynamic services under the Internet of Things environment is proposed. According to the probability that the allele on the fragment to be vaccinated has appeared in the memory bank, the detection strategy is used to judge whether the solution is illegal. By comparing other values that have appeared on the gene locus, the optimal individual can be formed to judge whether the optimal gene or the local optimality. And through simulated annealing to inoculate with probability, the resource allocation and optimization for the dynamic service of the agricultural Internet of Things are solved. Evaluate the total service cost and total service time under the two service provision strategies. The results confirmed the superiority of the proposed improved AIGE algorithm. This article proposes a service cost and service time minimization model from the perspective of service providers. In the future, the cost of the service request party will be considered to make the Internet of Things service model more perfect.

## Data Availability

The data included in this paper are available without any restriction.

## Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

## Authors' Contributions

The main idea of this paper is proposed by Haihong Liang. The algorithm design and experimental environment construction are completed by Haihong Liang. The experimental verification was completed by Haihong Liang. The writing of the article is completed by Haihong Liang. And the writing guidance, English polish, and funding of the project are completed by Haihong Liang.

## Acknowledgments

This work was supported by the Zhejiang Philosophy and Social Science Foundation Project (no. 15NDJC290YBM) and the Fund Project of the Science and Technology Department of Zhejiang Province (no. 2015C35040).

## References

- [1] P. Ray, "A survey on Internet of Things architectures," *Journal of King Saud University-Computer and Information Sciences*, vol. 30, no. 3, pp. 291–319, 2018.
- [2] A. Tzounis, N. Katsoulas, T. Bartzanas, and C. Kittas, "Internet of Things in agriculture, recent advances and future challenges," *Biosystems Engineering*, vol. 164, pp. 31–48, 2017.
- [3] M. Conti, A. Dehghantaha, K. Franke, and S. Watson, "Internet of Things security and forensics: challenges and opportunities," *Future Generation Computer Systems*, vol. 20, no. 1, pp. 20–34, 2018.
- [4] E. Sisinni, A. Saifullah, S. Han, U. Jennehag, and M. Gidlund, "Industrial Internet of Things: challenges, opportunities, and directions," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 11, pp. 4724–4734, 2018.
- [5] O. Elijah, T. A. Rahman, I. Orikumhi, C. Y. Leow, and M. H. D. N. Hindia, "An overview of Internet of Things (IoT) and data analytics in agriculture: benefits and challenges," *IEEE Internet of Things Journal*, vol. 5, no. 5, pp. 3758–3773, 2018.
- [6] K. L. Krishna, O. Silver, W. F. Malende, and K. Anuradha, "Internet of Things application for implementation of smart agriculture system," in *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)*, pp. 54–59, Palladam, India, February 2017.
- [7] F. J. Ferrández-Pastor, J. M. García-Chamizo, M. Nieto-Hidalgo, and J. Mora-Martínez, "Precision agriculture design method using a distributed computing architecture on Internet of Things context," *Sensors*, vol. 18, no. 6, pp. 17–31, 2018.
- [8] I. Mat, M. R. M. Kassim, A. N. Harun, and I. M. Yusoff, "Smart agriculture using Internet of Things," in *2018 IEEE Conference on Open Systems (ICOS)*, pp. 54–59, Langkawi, Malaysia, November 2018.
- [9] K. Leng, L. Jin, W. Shi, and I. van Nieuwenhuysse, "Research on agricultural products supply chain inspection system based on Internet of Things," *Cluster Computing*, vol. 22, Supplement 4, pp. 8919–8927, 2019.
- [10] X. Feng, F. Yan, and X. Liu, "Study of wireless communication technologies on Internet of Things for precision agriculture," *Wireless Personal Communications*, vol. 108, no. 3, pp. 1785–1802, 2019.
- [11] M. Ge, H. Bangui, and B. Buhnova, "Big data for Internet of Things: a survey," *Future Generation Computer Systems*, vol. 87, pp. 601–614, 2018.



- [12] A. Khanna and S. Kaur, "Evolution of Internet of Things (IoT) and its significant impact in the field of precision agriculture," *Computers and Electronics in Agriculture*, vol. 157, pp. 218–231, 2019.
- [13] I. U. Din, S. Hassan, A. Almogren, F. Ayub, and M. Guizani, "PUC: packet update caching for energy efficient IoT-based information-centric networking," *Future Generation Computer Systems*, vol. 11, no. 1, pp. 634–643, 2020.
- [14] B. B. Gupta and M. Quamara, "An overview of Internet of Things (IoT): architectural aspects, challenges, and protocols," *Concurrency and Computation: Practice and Experience*, vol. 32, no. 21, pp. 40–49, 2020.
- [15] S. Popli, R. Jha, and S. Jain, "A survey on energy efficient narrowband Internet of Things: architecture, application and challenges," *IEEE Access*, vol. 7, no. 9, pp. 16739–16776, 2018.
- [16] K. Jha, A. Doshi, P. Patel, and M. Shah, "A comprehensive review on automation in agriculture using artificial intelligence," *Artificial Intelligence in Agriculture*, vol. 2, no. 4, pp. 1–12, 2019.
- [17] D. I. Patrício and R. Rieder, "Computer vision and artificial intelligence in precision agriculture for grain crops: a systematic review," *Computers and Electronics in Agriculture*, vol. 15, no. 3, pp. 69–81, 2018.
- [18] E. Elahi, C. Weijun, H. Zhang, and M. Nazeer, "Agricultural intensification and damages to human health in relation to agrochemicals: application of artificial intelligence," *Land Use Policy*, vol. 8, no. 21, pp. 461–474, 2019.
- [19] S. Liu, L. Guo, H. Webb, X. Ya, and X. Chang, "Internet of Things monitoring system of modern eco-agriculture based on cloud computing," *IEEE Access*, vol. 7, pp. 37050–37058, 2019.
- [20] N. Ahmed, D. De, and I. Hussain, "Internet of Things (IoT) for smart precision agriculture and farming in rural areas," *IEEE Internet of Things Journal*, vol. 5, no. 6, pp. 4890–4899, 2018.
- [21] J. T. Zhang, H. J. Huang, and X. Wang, "Resource provision algorithms in cloud computing: a survey," *Journal of Network and Computer Applications*, vol. 64, pp. 23–42, 2016.
- [22] Z. Cai, X. Li, and J. Gupta, "Heuristics for provisioning services to workflows in XaaS clouds," *IEEE Transactions on Services Computing*, vol. 9, no. 2, pp. 250–263, 2016.
- [23] N. Ghosh, S. K. Ghosh, and S. K. Das, "SelCSP: a framework to facilitate selection of cloud service providers," *IEEE Transactions on Cloud Computing*, vol. 3, no. 1, pp. 66–79, 2015.
- [24] J. Ding, R. Yu, Y. Zhang, S. Gjessing, and D. H. K. Tsang, "Service provider competition and cooperation in cloud-based software defined wireless networks," *IEEE Communications Magazine*, vol. 53, no. 11, pp. 134–140, 2015.
- [25] Z. Zhu, X. Zhang, M. Li, and X. Liu, "Evolutionary multi-objective workflow scheduling in cloud," *IEEE Transactions on Parallel and Distributed Systems*, vol. 27, no. 5, pp. 1344–1357, 2016.
- [26] S. H. Madni, "Resource scheduling for infrastructure as a service (IaaS) in cloud computing: challenges and opportunities," *Journal of Network and Computer Applications*, vol. 68, no. 12, pp. 173–200, 2016.
- [27] S. Zhao, L. Yu, and B. Cheng, "An event-driven service provisioning mechanism for IoT (Internet of Things) system interaction," *IEEE Access*, vol. 4, no. 1, pp. 5038–5051, 2016.
- [28] A. Romay, S. Kohlbrecher, A. Stumpf et al., "Collaborative autonomy between high-level behaviors and human operators for remote manipulation tasks using different humanoid robots," *Journal of Field Robotics*, vol. 34, no. 2, pp. 333–358, 2017.
- [29] Y. Li, Y. Huang, M. Zhang, and L. Rajabion, "Service selection mechanisms in the Internet of Things: a systematic and comprehensive study," *Cluster Computing*, vol. 54, no. 9, pp. 1–21, 2019.
- [30] A. Klein, F. Ishikawa, and S. Honiden, "SanGA: a self-adaptive network-aware approach to service composition," *IEEE Transactions on Services Computing*, vol. 7, no. 3, pp. 452–464, 2014.
- [31] Q. Z. Sheng, B. Benatallah, Z. Maamar, and A. H. H. Ngu, "Configurable composition and adaptive provisioning of web services," *IEEE Transactions on Services Computing*, vol. 2, no. 1, pp. 34–49, 2009.

## Research Article

# CPEH: A Clustering Protocol for the Energy Harvesting Wireless Sensor Networks

Yu Han <sup>1</sup>, Jian Su <sup>2</sup>, Guangjun Wen <sup>1</sup>, Yiran He <sup>1</sup> and Jian Li <sup>1</sup>

<sup>1</sup>University of Electronic Science and Technology of China, 611731, China

<sup>2</sup>Nanjing University of Information Science & Technology, 210044, China

Correspondence should be addressed to Jian Li; lj001@uestc.edu.cn

Received 10 February 2021; Revised 4 March 2021; Accepted 26 March 2021; Published 12 April 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Yu Han et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In the last decade, energy harvesting wireless sensor network (EHWSN) has been well developed. By harvesting energy from the surrounding environment, sensors in EHWSN remove the energy constraint and have an unlimited lifetime in theory. The long-lasting character makes EHWSN suitable for Industry 4.0 applications that usually need sensors to monitor the machine state and detect errors continuously. Most wireless sensor network protocols have become inefficient in EHWSN due to neglecting the energy harvesting property. In this paper, we propose CPEH, which is a clustering protocol specially designed for the EHWSN. CPEH considers the diversity of the energy harvesting ability among sensors in both cluster formation and intercluster communication. It takes the node's information such as local energy state, local density, and remote degree into account and uses fuzzy logic to conduct the cluster head selection and cluster size allocation. Meanwhile, the Ant Colony Optimization (ACO) as a reinforcement learning strategy is utilized by CPEH to discover a highly efficient intercluster routing between cluster heads and the base station. Furthermore, to avoid cluster dormancy, CPEH introduces the Cluster Head Relay (CHR) strategy to allow the proper cluster member to undertake the cluster head that is energy depletion. We make a detailed simulation of CPEH with some famous clustering protocols under different network scenarios. The result shows that CPEH can effectively improve the network throughput and delivery ratio than others as well as successfully solve the cluster dormancy problem.

## 1. Introduction

The automatic and unmanned operation is one of the most prominent characters in the Industry 4.0 era [1, 2]. To monitor the industrial system's state and detect errors, the Industry 4.0 applications need a large number of sensors attached to machines to report their sensing data to the control center periodically [3]. Traditional wireless sensor networks (WSN) face the weakness of energy supply [4]. When the sensors run out of their energy, the network will suffer the performance wreck. Replacing the sensor's battery will bring extra cost and even be impossible in some harsh scenarios. There are currently many protocols designed for WSN trying to extend the practical network lifetime [5–8] by saving and balancing the energy consumption among sensors. However, the innate defect of WSN cannot be covered up. Since the sense is a long-term task, traditional WSN will no longer be suitable in the Industry 4.0 era. In the last decade, the rapid develop-

ment of energy harvesting technology [9] makes tiny commercial sensors with energy harvesting ability available. Researchers proposed the energy harvesting wireless sensor network (EHWSN) to realize the long-term operation [10, 11]. Sensors in EHWSN harvest energy from their surrounding environments and utilize the harvested energy to power themselves. EHWSN eliminates the battery requirement and has an unlimited lifetime, in theory, making it a promising technology in Industry 4.0 era.

EHWSN typically consists of a base station (BS) and a larger number of sensors. Sensors generate their sensing data shortly and periodically to guarantee the monitoring accuracy and quickly respond to the error. EHWSN shall handle the heavy generated data timely and transmit all the information to the BS, which is high energy costing. However, in the majority of scenarios, the energy harvesting rates of sensors are limited and individual [12]. To improve the network performance, effective routing protocols are necessary for

EHWSN to cut down the energy consumption while better utilizing the harvested energy [13]. As a hierarchy approach, clustering has been proved to have the inherited advantages on energy efficiency and scalability compared with flat routing protocols [14, 15], making it suitable for EHWSN.

Clustering protocols work in rounds. Every round, the cluster head (CH) will collect the data from its cluster members (CMs) and upload the handled data to the BS. Since CMs are usually close to each other, the data they generate may have high correlations. The CH will take local data fusion to remove the redundancy, thus reducing the data quantity. The protocol can balance the workload over different sensors and utilize the network harvested energy properly by reconstructing the clusters, whereas despite the advantages, at the present stage, most clustering protocols are designed for traditional WSN. There will be some challenges for them to work in the EHWSN condition. First, the primary design purpose of current clustering protocols is to extend the lifetime of WSN. Nevertheless, the leading aspiration of EHWSN has turned to maximize the network performance under the energy harvesting restrictions. Second, both the cluster formation and the intercluster routing of current clustering protocols do not consider the different energy harvesting rates between sensors, resulting in a performance decline. Last, in EHWSN, the sensor may deplete the energy and return to the sleep state when serving as a CH. Current clustering protocols cannot respond to this condition, causing the cluster dormancy phenomenon.

In this paper, we propose CPEH, which is a clustering protocol designed explicitly for EHWSN. CPEH cares about the diversity of the energy harvesting ability among sensors in both cluster formation and intercluster communication. In the cluster formation process, CPEH considers the sensor's local energy state, local node density, and remote degree. It then uses fuzzy logic to conduct the cluster head competition distributedly. In the intercluster routing discovery, the Ant Colony Optimization (ACO) is utilized to help the BS discover the proper paths for different CHs. To avoid cluster dormancy, CPEH introduces the Cluster Head Relay (CHR) strategy, which will choose the appropriate CM to undertake the sleep CH. Compared with other clustering protocols, the simulation result proves that CPEH will effectively improve the network delivery ratio and elevate the network throughput. Meanwhile, the CHR strategy can successfully maintain the cluster working normally when the initial CH runs out of its energy.

The rest of this paper is organized as follows. Section 2 gives a brief review of current clustering protocols. In Section 3, both the network model and the energy model used in this paper are described. In Section 4, we introduce our proposed protocol CPEH in detail. The simulation result is discussed in Section 5. Finally, Section 6 concludes the paper and introduces the further work.

## 2. Related Works

Clustering protocols have been well studied in the last decade. The Low Energy Adaptive Clustering Hierarchy (LEACH) [16] protocol is the most famous clustering proto-

col for periodical sensing applications. LEACH works in rounds and uses a random rotation to select the CH in order to balance the energy dissipation between sensors. The rotation guarantees each sensor to become a CH once and only once in every  $1/P$  round, where  $P$  is the expected cluster head ratio. Every CH adopts a local data aggregation to decrease the data amount sending to the BS. LEACH achieves a better performance than traditional flat routing protocols. However, researchers in [17] point out that the selection of CHs in LEACH does not consider the sensor's energy state. The low energy node will die rapidly, once it becomes a CH, limiting the network lifetime. To solve that, they propose an Energy-Efficient Clustering Scheme (EECS), which uses the residual energy as the concerned parameter of the CH selection. The candidate CHs with the most energy within the competitive radius will become the final CH. To balance the energy consumption, CMs in EECS choose their CHs based on a cost function that gives the CHs near the BS a higher priority. Hence, the energy consumption between CHs is balanced. Nevertheless, both LEACH and EECS are based on single-hop intercluster communication, causing them to be inefficient in large-scale networks due to long-range transmission. Hybrid Energy-Efficient Distributed Clustering (HEED) [18] employs an iteration strategy to construct the cluster and carry out a rarely uniform CH distribution. The communication between CHs and the BS is multihop to reduce the transmission cost. HEED achieves a good performance. However, the iteration process will increase the energy overhead caused by the exchanged control packets. Researchers in [19] propose the Unequal Cluster-based Routing (UCR) protocol which groups the nodes into clusters of unequal sizes to strike the energy dissipation imbalance between different CHs. The clusters near the BS will have smaller sizes in UCR so that the CHs can achieve the energy trade-off between the intracluster and intercluster.

It has been proved that the CH selection is typically an NP hard problem. Recently, with the fast development in computational intelligence, many researchers try to use the metaheuristic approach or fuzzy logic technology to resolve the CH selection. In [20], the BS uses particle swarm optimization (PSO) to optimize the set of CHs. The fitness function of PSO tries to decrease the maximum average intracluster distance in the meanwhile keeping the CHs with more energy. The CH in the Clustering Protocol based on the Metaheuristic Approach (CMPA) [21] is selected based on the Harmony Search (HS) algorithm. The selection focuses on reducing energy consumption in the meanwhile keeping a balanced energy distribution. CPMA also creatively introduces the Artificial Bee Colony (ABC) algorithm to optimize the protocol parameters, making it robust to various application scenarios. Clustering protocols using the metaheuristic approach are centralized and have a relatively high computational power requirement for the BS. For the large-scale network with strict scalability demand, it is a better choice to use fuzzy logic to handle the cluster head competition distributedly. In [22], the Fuzzy Energy-Aware Unequal Clustering (EAUCF) algorithm uses fuzzy logic to guide the clustering process. The fuzzy system considers both the residual energy and the distance to the BS. It is aimed at balancing the energy

consumption between CHs by assigning the CHs with different competitive radius. In [23], the researcher proposes a Multiobjective Fuzzy Clustering Protocol (MOFCA). MOFCA takes the sensors' nonuniform distribution and movement into account, making it suitable for both stationary and evolving networks. MOFCA has the same fuzzy system output as EAUCF. On the contrary, MOFCA additionally adds the sensor's local density to the fuzzy system input. The consideration of local density can effectively decrease the intracluster communication cost and maintain robustness over various network distribution and movement conditions, whereas like most clustering protocols, MOCFA does not care about the intercluster routing. The improper routing between CHs may downgrade the performance, especially in large-scale networks.

All the clustering protocols introduced above are designed for the traditional WSN. Currently, there are only several clustering protocols for EHWSN. In [24], the researchers propose the Energy Neutral Clustering (ENC) protocol to provide perpetual network operation. ENC employs a novel Cluster Head Group (CHG) mechanism that allows a cluster to use multiple cluster heads to share heavy traffic load. In every round, the sensors in the CHG will take the role of CH in turns to make sure the data sent by CMs can always be received. ENC uses convex optimization to choose the best combination of cluster number and the CHG number. The simulation shows ENC can improve the throughput than others. However, owing to the CHG mechanism, the areas in which CHG nodes locate may not be monitored correctly. A novel centralized clustering protocol for EHWSN is proposed in [25]. The BS in [25] runs a modified discrete particle swarm optimization (DPSO) algorithm to select CHs for the networks. The optimization considers both the transmission cost and the sensor's residual energy but ignores the difference of the energy harvesting ability between sensors, leading to a performance discant. Researchers in [26] pointed out that the energy harvesting process usually does not match the real energy demand, and sensors suffer from occasional energy shortages, mostly when they serve as the CH. To address this issue, they propose a Distance-and-Energy-Aware Routing with Energy Reservation (DEARER). DEARER encourages the sensor with more residual energy and a short distance to the sink more likely to become the CH. It also allows sensors to save their harvested energy for future use. The simulation results show that DEARER can achieve good performance. However, the complexity of DEARER is much higher than others, which causes it hard to implement.

### 3. System Models

**3.1. Network Model.** As shown in Figure 1, the EHWSN in this paper utilizes the CPEH, a multihop intercluster routing-based clustering protocol, to coordinate its network communication. The network assumptions are summarized as follows:

- (1) The BS is energy unlimited and always has sufficient knowledge of the entire network. Sensors are ran-

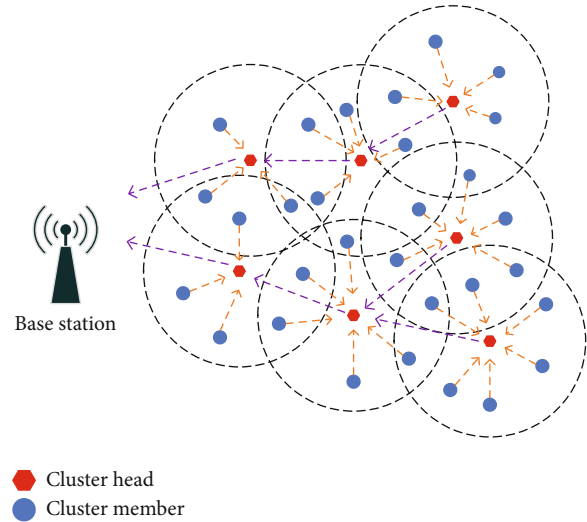


FIGURE 1: The primary network structure of the CPEH. CPEH is a multihop intercluster communication-based clustering protocol.

domly placed throughout the entire network. All the sensors have the same initial energy and maximum energy capacity. However, the energy harvesting rates among sensors are different. Both the BS and the sensor are motionless once deployed

- (2) All the sensors can communicate with the BS directly and have the ability to adjust their transmitting power according to the distance
- (3) All the network links are symmetric. Sensors can estimate the communication distance based on the received signal power
- (4) The network is homogeneous. The data generated by sensors in the same cluster have a relatively high correlation. The cluster head can make the local data aggregation to reduce the redundancy. The aggregation ratio is the same in different clusters

**3.2. Energy Model.** For energy harvesting, we assume that sensors can continuously harvest energy from their surrounding environments at a fixed speed. However, due to the diversity of location and hardware, the energy harvesting rates differ between sensors. *Since sensors usually harvest their energy from the heat or the vibration of their associated machines, we believe that this assumption is reasonable for the EHWSN used in the Industry 4.0 applications.* We use the *harvest-use* model to describe the energy availability of sensors. The harvested energy can be used immediately, which reduces the complexity of the energy management and easy to be implemented.

As shown in Figure 2, we employ the classical radio energy dissipation model used in [16]. The transmitter will consume energy to run the radio electronics and the power amplifier, whereas the receiver only dissipates energy to run the radio electronics. Represent the packet size as  $l$  bits and the communication distance as  $d$ ; the energy cost of

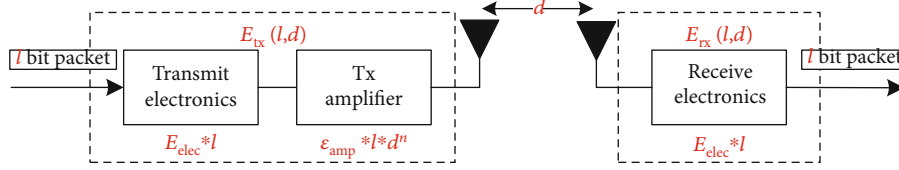


FIGURE 2: The radio energy dissipation model used in this paper.

transmitting ( $E_{tx}$ ) and receiving ( $E_{rx}$ ) a packet can be expressed as follows:

$$E_{tx}(l, d) = \begin{cases} E_{elec} \cdot l + E_{fs} \cdot d^2 \cdot l, & d \leq d_0, \\ E_{elec} \cdot l + E_{mp} \cdot d^4 \cdot l, & d > d_0, \end{cases} \quad (1)$$

$$E_{rx}(l, d) = E_{elec} \cdot l, \quad (2)$$

where  $E_{elec}$  is the energy cost per bit by the radio electronics of both transmitter and receiver and  $E_{fs} \cdot d^2$  and  $E_{mp} \cdot d^4$  represent the energy needed per bit for the power amplifier under the free space channel ( $d^2$ ) and the multipath channel ( $d^4$ ).  $d_0$  is the threshold distance defined as

$$d_0 = \sqrt{\frac{E_{fs}}{E_{mp}}}. \quad (3)$$

Notice that  $E_{elec}$ ,  $E_{fs}$ , and  $E_{mp}$  depend on the transceiver characters and the acceptable bit-error rate [19]. They should be set carefully. For data aggregation, assuming the CH will consume  $E_{da}$  energy to handle one-bit data and the cluster has  $W$  sensors (including the CH), the energy dissipation for the CH in the aggregation process is

$$E_{agg} = l \cdot W \cdot E_{da}. \quad (4)$$

## 4. CPEH Introduction

**4.1. CPEH Overall.** Figure 3 exhibits the basic communication timeline of CPEH, which is operated in rounds. Each round is comprised of a setup phase and a working phase. To better utilize the harvested energy and avoid some particular sensors' overuse, the CHs should be selected reasonably. Hence, awake sensors shall have different chances to be the CH based on the considerations of their energy states, locations, and surrounding conditions. The same considerations should also be taken to assign the CHs with different allowed sizes. At the beginning of the setup phase, every awake sensor will proceed with the cluster head election distributedly based on fuzzy logic, which we will give a detailed description in the following paper. After that, regular sensors will choose the CH to join based on the distance and size of the CH. Then, the CH will generate a TDMA schedule and broadcast it to all its members. At the end of the setup phase, the BS will use the ACO algorithm to discover the best inter-cluster route for each CH and broadcast the result to all the CHs.

Owing to the exchanging of control messages, the setup phase is relatively energy costing. To reduce the communication overhead, CPEH combines several frames into one working phase. In each frame, based on the TDMA schedule, the awake CM will transmit the sensing data to the CH in its associated slot and keep silent in others to save energy and avoid the collision. Once the CH detects the current energy lower than the sleep threshold, it will execute the CHR strategy to find the replacer and broadcast the result to all the awake CMs in the particular slot (Slot CHR). The chosen CM will become the successive CH in the following frames. Notice that if the network burden is heavy, the CHR strategy may be executed several times to maintain the cluster usually. After receiving the data from all the awake CMs, the CH will first aggregate the data and then upload the fusion data to the BS through the multihop routing.

The direct-sequence spread spectrum (DSSS) is utilized to avoid intercluster collision. CPEH assigns a universal spreading code and a unique spreading code to every sensor. The universal spreading code is used to exchange control messages in the setup phase and route packets between CHs in the working phase. On the contrary, once a sensor becomes a CH, it will confirm all the members using its unique spreading code for the intracluster communication. Hence, the adjacent clusters will have different spreading codes, eliminating the collisions among clusters.

Since the harvested energy may not support the continuous work, sensors under CPEH will operate in a sleep-wake mode. Define  $E_{capacity}$  as the sensor's energy storage capacity,  $E_{sleep}$  as the sensor's sleep threshold, and  $E_{awake}$  as the sensor's awake threshold. Then, we have  $E_{sleep} = 0.1 \cdot E_{capacity}$  and  $E_{awake} = 0.2 \cdot E_{capacity}$ . If the sensor detects that its residual energy has dropped off  $E_{sleep}$ , it turns to the sleep state to maintain the basic procedure and recharge. Once the energy reaches  $E_{awake}$ , the dormant sensor will wake up again and work normally.

### 4.2. CPEH Detailed

**4.2.1. Cluster Construction.** The cluster construction of CPEH is based on fuzzy logic. At the beginning of the setup phase, each awake sensor will broadcast a *State\_Message*, which contains its current residual energy and the energy harvesting rate, in the radius  $R$ . Based on the *State\_Message*, awake sensors will calculate the average residual and the average harvesting rate within  $R$ , which will be further used. Then, awake sensors will use the fuzzy logic system to estimate their chances to become the CHs and the allocated cluster sizes. The complicated fuzzy logic system working process is described as follows.

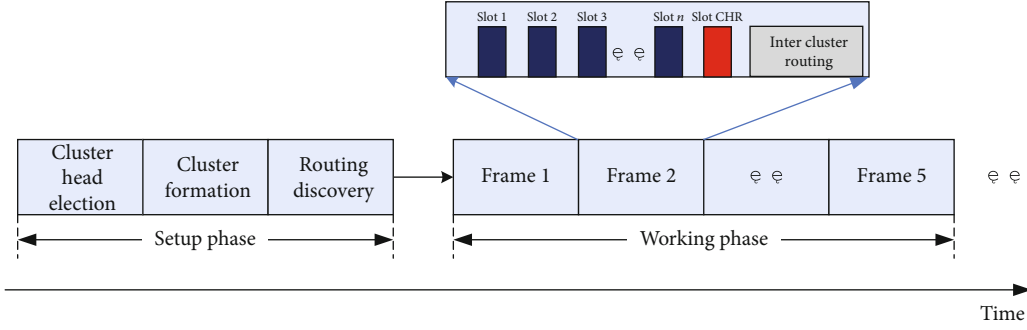


FIGURE 3: The basic communication timeline of CPEH. CPEH works in rounds. Every round consists of a setup phase and a working phase.

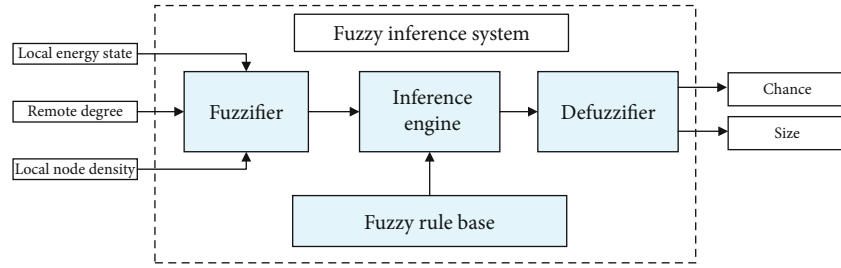


FIGURE 4: The structure of the fuzzy logic system. Sensor inputs the local energy state, remote degree, and the local node density. The system outputs the chance and the size.

The fuzzy logic system of CPEH is shown in Figure 4. It considers three different inputs: the local energy state, the local node density, and the distance to the BS.

(1) *Local Energy State*. The local energy state of sensor  $i$  is defined as

$$S_{i_{es}} = \left( \frac{S_{i_{re}}}{S_{i_{energy\_ave}}} \right)^\alpha \cdot \left( \frac{S_{i_{har}}}{S_{i_{har\_ave}}} \right)^\beta, \quad (5)$$

where  $S_{i_{re}}$ ,  $S_{i_{har}}$  means the residual energy and the energy harvesting rate of the sensor  $i$  and  $S_{i_{energy\_ave}}$ ,  $S_{i_{har\_ave}}$  represents the average residual energy and the average energy harvesting rate of all the awake sensors in the radius  $R$  of  $i$ .  $\alpha$  and  $\beta$  are the impact factors; we choose  $\alpha = 2$  and  $\beta = 1$  in this paper. The larger the local energy state is, the larger the chance and the size should be so that the sensor can fully exploit the harvested energy to undertake more works. The corresponding fuzzy linguistic variables are Good, Moderate, and Bad. Figure 5 shows the membership function of those linguistic variables.

(2) *Local Node Density*. The local node density of sensor  $i$  is defined as

$$S_{i_{den}} = \frac{\text{network\_area} \cdot N_{i_{local}}}{\pi \cdot R^2 \cdot N_{total}}, \quad (6)$$

where  $\text{network\_area}$  means the area of the entire network,  $N_{total}$  means network sensor number, and  $N_{i_{local}}$  means the

number of total awake sensors in the radius of sensor  $i$ . We believe that the fuzzy logic system should assign the sensor, which has a large density, with a higher chance and a bigger size so that the entire intracluster communication cost can be reduced. The corresponding fuzzy linguistic variables are Sparse, Normal, and Dense. Figure 6 shows the membership function of those linguistic variables.

(3) *Remote Degree*. Let  $d_{i\_to\_BS}$  means the distance between sensor  $i$  and the BS. Let  $d_{max}$  mean the max  $d_{i\_to\_BS}$  among all the sensors. We define

$$d_{i\_remote} = \frac{d_{i\_to\_BS}}{d_{max}}, \quad 0 < d_{i\_remote} \leq 1 \quad (7)$$

as the parameter to show the sensor's relative remote degree. The fuzzy logic system should give the CH nearing the BS a smaller cluster size to make the trade-off between intercluster and intracluster energy consumption. The corresponding fuzzy linguistic variables are Far, Medium, and Near. Figure 7 shows the membership function of those linguistic variables.

The fuzzy logic system has two outputs: chance and size. We give the chance nine fuzzy linguistic variables: Very Low, Low, Rather Low, Low Medium, Medium, High Medium, Rather High, High, and Very High. The corresponding membership functions are shown in Figure 8. The fuzzy linguistic variables of size are Very Small, Small, Rather Small, Low Medium, Medium, High Medium, Rather Large, Large, and

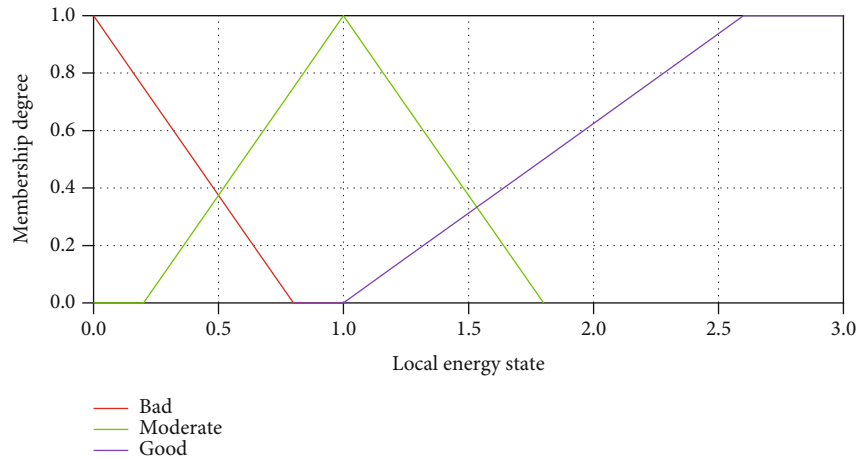


FIGURE 5: Membership function for local energy state.

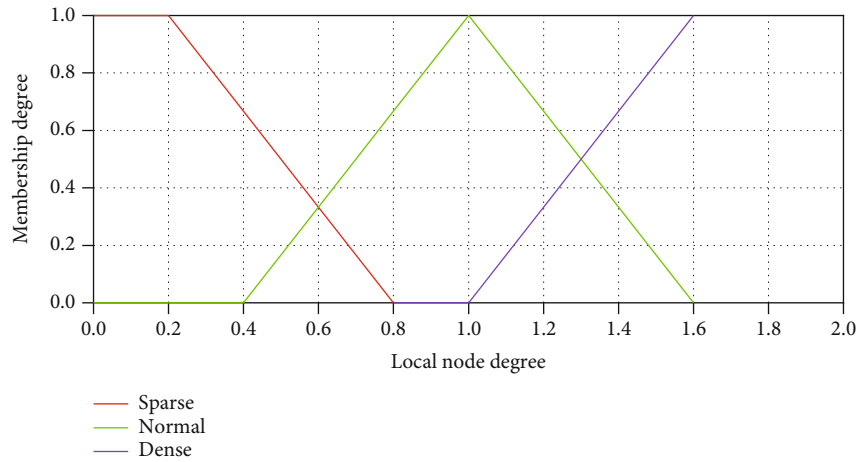


FIGURE 6: Membership function for local node degree.

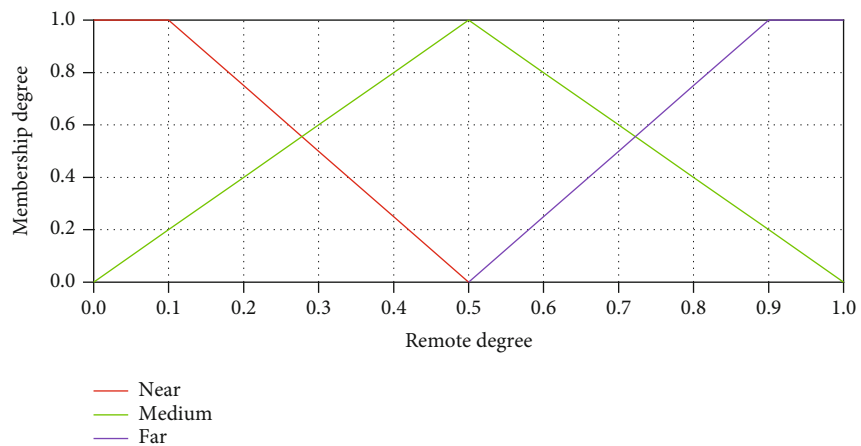


FIGURE 7: Membership function for remote degree.

Very Large. Figure 9 shows the membership function of those linguistic variables.

Once the sensor inputs the crisp values into the fuzzy logic system, the fuzzifier will transform those values to the

associated linguistic variables based on the membership functions. Then, the Mamdani [27] fuzzy inference engine will make the inference according to the input linguistic variables based on the fuzzy rules shown in Table 1. After that,

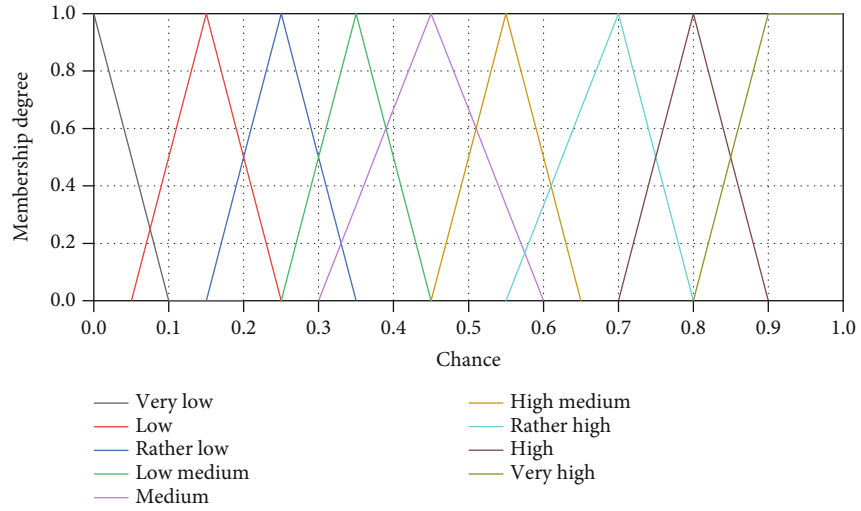


FIGURE 8: Membership function for remote degree.

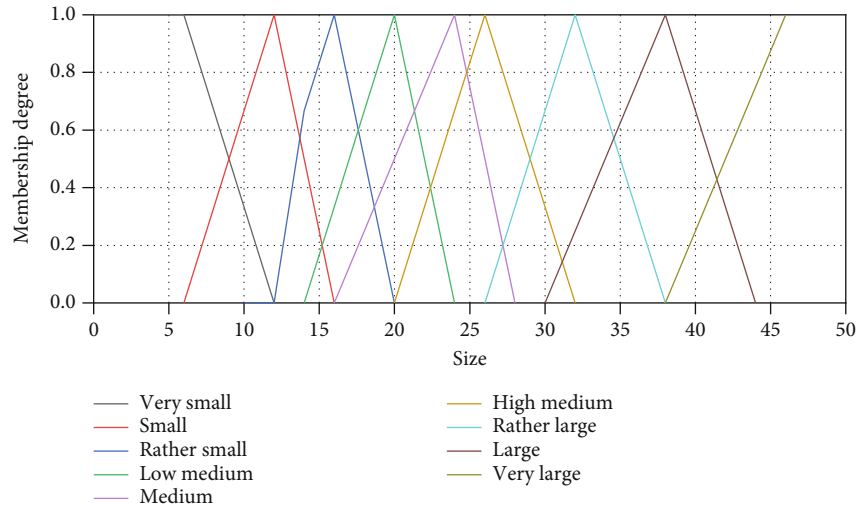


FIGURE 9: Membership function for size.

TABLE 1: The detailed fuzzy rules. The fuzzy rule base has 27 different rules.

No.	Input variables			Output variables	
	Local energy state	Remote degree	Local node density	Chance	Size
1	Bad	Near	Sparse	Very Low	Very Small
2	Bad	Near	Normal	Low	Very Small
3	Bad	Near	Dense	Rather Low	Small
13	Moderate	Medium	Sparse	Low Medium	Low Medium
14	Moderate	Medium	Normal	Medium	Medium
15	Moderate	Medium	Dense	High Medium	High Medium
25	Good	Far	Sparse	Rather High	Rather Large
26	Good	Far	Normal	High	Large
27	Good	Far	Dense	Very High	Very Large

the inferred linguistic variables will be defuzzified based on the center of area method. The fuzzy logic system finally outputs the crisp value of the chance and size.

After the fuzzy inference, the awake sensor will then broadcast a *Competition\_Message* containing its chance and ID, and it will record the corresponding information of its



```

Cluster Construction Process for sensor  $i$ 
1: At the beginning of the Round
2: If  $S_i.state = awake$  then
3:   Broadcast a State_Message within  $R$ 
4: End if
5: On receiving the State_Message from its neighbor  $j$ 
6:   add  $j$  to its neighbor set  $S_i.neighbor$ 
7:   calculate its local energy state and local node density
8:   infer its chance and size base on the fuzzy logic system
9: Broadcast a Competition_Message ( $S_i.chance, S_i.ID$ )
10: While the cluster head election time has not expired, do
11:   if  $S_i.chance > S_j.chance, \forall S_j \in S_i.neighbor$  then
12:     broadcast a Winner_Message ( $S_i.ID$ ) and then EXIT
13:   end if
14:   on receiving a Winner_Message from sensor  $j$ 
15:   if  $S_j \in S_i.neighbor$  then
16:     broadcast a Quit_Message ( $S_i.ID$ ) and then EXIT
17:   end if
18:   on receiving a Quit_Message from sensor  $j$ 
19:   if  $S_j \in S_i.neighbor$  then
20:     remove sensor  $j$  from  $S_i.neighbor$ 
21:   end if
22: end while
23: If  $S_i.role = CH$  then
24:   On receiving the Join_Message from sensor  $j$ 
25:   if the current cluster size less than  $S_i.size$ 
26:     Accept the sensor  $j$ , sending back the Accept_Message
27:   else Sending back the Reject_Message
28:   end if
29: Else
30:   Sending the Join_Message to the nearest CH
31:   On receiving the Accept_Message, finish
32:   On receiving the Reject_Message, delete the CH and repeat 30
33:   if all the neighbor CHs refuse to accept sensor  $i$ 
34:      $S_i.role = CH$ 
35:   end if
36: End if

```

PSEUDOCODE 1: Pseudocode of the cluster construction process in CPEH.

neighbors based on the received *Competition\_Message*. After that, the awake sensor will broadcast a *Winner\_Message* and become the final CH if it has the largest chance among its neighbors. If an awake sensor receives a *Winner\_Message* from its neighbors, it will quit the election and broadcast a *Quit\_Message*. The awake sensor will remove the sensor from its neighbors if it receives the corresponding *Quit\_Message*.

The non-CH sensors may receive more than one *Winner\_Message*. It will then choose the nearest CH to join and send a *Join\_Message* to the CH. Once the CH receives a *Join\_Message*, it will accept the sensor by sending back an *Accept\_Message* if the cluster is not full. Otherwise, a *Reject\_Message* will be sent to the sensor. The sensor will send another *Join\_Message* to the next nearest CH on receiving the *Reject\_Message* and may repeat this process until successfully joining a cluster. There may be the condition that all the clusters have no room for a sensor to join. In this case, the sensor becomes a CH by itself. Our simulation shows this condition is quite rare. We summarize the pseudocode of the cluster construction in Pseudocode 1.

After the cluster construction, the CH will broadcast a TDMA schedule that assigns the slots for its members. In the working phase, sensors will transmit their data in their associated slot and keep silent in others to save energy and avoid the collision. CPEH combines several frames into one round. Hence, awake sensors may deplete their energy and turn to sleep in the middle of the current round. If the depleted sensor is the CM, it will just give up its slot, whereas if the depleted sensor is the CH, the CHR strategy will be executed, which we will discuss later. Notice the sleep sensors at the setup phase may wake up halfway in the working round. However, those sensors will keep silent until the round ending since there is no slot designed.

Since the cluster construction is conducted fully distributedly by sensors, we analyze the control message complexity, reflecting the overhead caused by the message exchanging. Assuming there are  $N$  awake sensors at the setup phase, every awake sensor will broadcast a *State\_Message* and a *Competition\_Message*. If  $M$  sensors become the CH, there will be  $M$  *Winner\_Message* and  $N - M$  *Quit\_Message*. Then,

the number of control messages in the cluster head election process is  $N + N + M + (N - M) = 3N$ . Let  $k_i$  represent the number of attempts of  $CM_i$  before it joins a cluster. The number of control messages in the cluster formation process can be represented as  $\sum_1^{N-M} 2k_i = 2(N - M)K_{\text{average}}$ , where  $K_{\text{average}}$  means the average attempt number and is tested to be less than 3 in all the simulations. Hence, the total control messages in the cluster construction are  $(3 + 2K_{\text{average}})N - 2MK_{\text{average}}$ . The corresponding complexity is  $O(N)$ . Note that the control message number of CPEH is slightly larger than some cluster protocols such as LEACH, EAUCF, and MOFCA. However, we believe that this sacrifice will construct more proper clusters and enhance the final network performance.

**4.2.2. Intercluster Routing Discovery.** An effective intercluster routing protocol can significantly decrease network energy consumption and improve network performance. However, most clustering protocols only pay attention to the cluster construction but neglect the impact of intercluster routing on the final system performance. The CPEH considers both the energy cost and the energy state in the routing discovery and decides the final result in the view of the entire network. As the basic principle of ACO is the positive feedback mechanism [28], ACO can be treated as a reinforcement learning approach. A larger number of researches have proved that ACO has a unique advantage in solving the routing problem over other metaheuristic algorithms.

The routing discovery is based on iterations, which is both energy costing and complex. We assign this work to the BS, which can reduce the overhead of CHs and get the result quickly. After the cluster construction, each CH will transmit a control packet to the BS, containing its current residual energy and energy harvesting rate. The BS will calculate the energy states of each CH  $i$ :  $CH_i.es = (CH_i.re/E_{\text{cap}}) \cdot (CH_i.har/E_{\text{harmax}})$  and the possible next-hop set of each CH  $i$ :  $Set_i = \{CH_0, CH_j\}$ ,  $d_{CH_i \rightarrow BS} > d_{CH_j \rightarrow BS}$ , where  $E_{\text{cap}}$  means the energy capacity and  $E_{\text{harmax}}$  means the max harvesting rate. We let  $CH_0$  represent the BS. Then, the BS will discover the intercluster routes based on the following steps. We take the route between  $CH_i$  and the BS as an example.

**Step 1 (route searching).** The BS place  $n$  ants in  $CH_i$  at a fixed time interval to find the path to the BS. Each ant will choose the next-hop  $CH_j$  according to the following equation:

$$P_{ij} = \frac{[\tau_{ij}(t)]^\theta [\eta_{ij}]^\nu}{\sum_{k \in Set_i} [\tau_{ik}(t)]^\theta [\eta_{ik}]^\nu}, \quad (8)$$

where  $P_{ij}$  represents the probability of the ant in  $CH_i$  choosing  $CH_j$  as its next hop.  $\tau_{ij}(t)$  is the pheromone trail density on the edge of  $(CH_i, CH_j)$  at the time  $t$ . The routing design should consider both the total energy consumption and the energy states of sensors on the path. Hence, we define the heuristic information  $\eta_{ij}$  as

$$\eta_{ij} = \frac{CH_j.es}{[\text{dis}(CH_i, CH_j)]^2}. \quad (9)$$

$\theta$  and  $\nu$  are impact factors. We give  $\theta = 1$  and  $\nu = 3$  to let the algorithm own a good ability in both local and global search. The ant will continue the searching until it reaches the BS.

**Step 2 (route evaluating).** When the ant reaches the BS, the BS will get the route information that can be represented by  $X = (x_1, x_2, x_3, \dots, x_m, x_{m+1})$ , where  $m$  is the number of CHs in the route,  $x_1$  means the source CH and  $x_{m+1}$  means the BS. Then, the BS will evaluate the quality of the route based on the objective function  $F$ :

$$F = Q \cdot \frac{E_{\text{ave}} \cdot E_{\text{min}}}{E(X)}, \quad (10)$$

where  $Q$  is the constant factor.  $E_{\text{ave}}$  and  $E_{\text{min}}$  are the average energy state and minimum energy state of the sensors on the route.  $E(X)$  is the total energy consumption of transmitting a data packet over the route and can be summarized as follows:

$$\begin{aligned} E(X) &= l \cdot \left( \sum E_{\text{receive}} + \sum E_{\text{transmit}} \right) \\ &= \left( E_{\text{elec}} \cdot (m - 1) + E_{\text{elec}} \cdot m \right. \\ &\quad \left. + \sum_{i=1}^m \begin{cases} E_{\text{fs}} \cdot d_{x_i, x_{i+1}}^2, & d_{x_i, x_{i+1}} \leq d_0 \\ E_{\text{mp}} \cdot d_{x_i, x_{i+1}}^4, & d_{x_i, x_{i+1}} > d_0 \end{cases} \right) \\ &= \left( E_{\text{elec}} \cdot (2m - 1) + \sum_{i=1}^m \begin{cases} E_{\text{fs}} \cdot d_{x_i, x_{i+1}}^2, & d_{x_i, x_{i+1}} \leq d_0 \\ E_{\text{mp}} \cdot d_{x_i, x_{i+1}}^4, & d_{x_i, x_{i+1}} > d_0 \end{cases} \right). \end{aligned} \quad (11)$$

$$\tau_{ij}(t+n) = (1 - \rho) \cdot \tau_{ij}(t) + \Delta\tau_{ij}, \quad \Delta\tau_{ij} = F_{\text{best}}. \quad (12)$$

**Step 3 (pheromone updating).** Based on the objective function, the BS can find the best route in one iteration. CPEH then uses the max-min ant system model [29] to update the pheromone. The max-min ant system model will only update the pheromone on the path of the best route. Hence, the convergence rate will be improved.

The pheromone is updated according to (12).  $\rho$  is the pheromone decay coefficient used to escape the local optima.  $F_{\text{best}}$  is the best objective function value in this iteration.

The BS will repeat those steps until the iteration time is reached. The best route information will be broadcasted to all the CHs at the end of the setup phase.

**4.2.3. The CHR Strategy.** The fuzzy logic system of CPEH is aimed at selecting the sensor with a good local energy state to become the CH. However, the heavy workload may still exhaust the CH's energy halfway in the working round, especially when the frame number is large. Most clustering protocols cannot handle this condition, causing the whole cluster

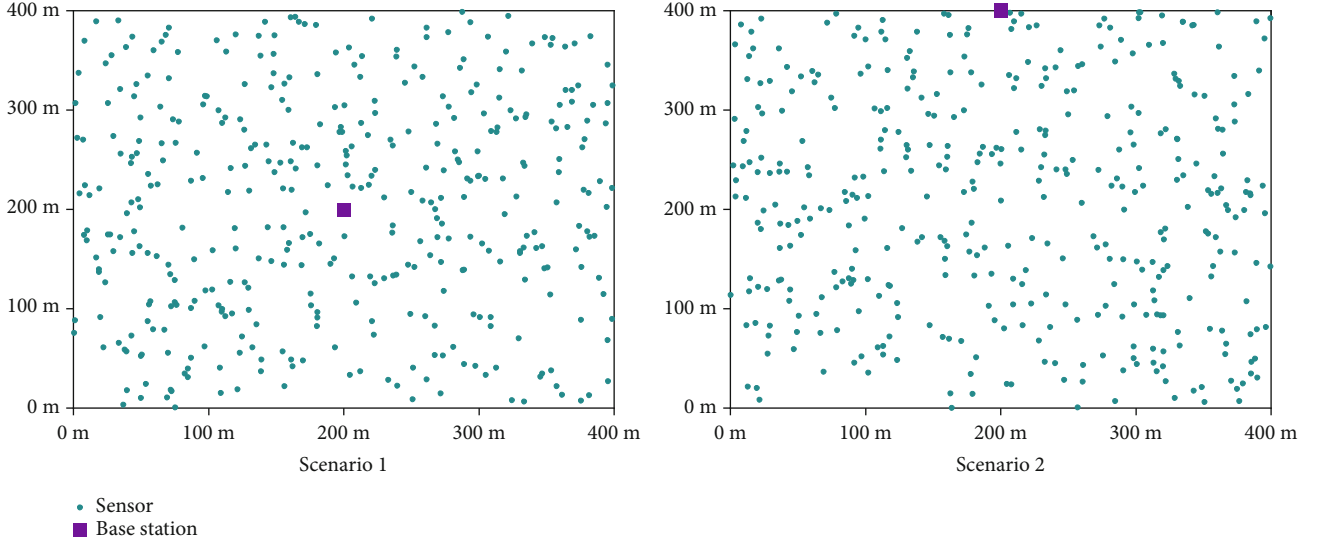


FIGURE 10: The sensors and BS deployment case in different scenarios.

to be silent until the end of the working round and wrecking the sensing accuracy. To solve this, CPEH utilizes the CHR strategy, a simple greedy approach, to help the depleted CH find the proper CM to undertake its job. Once the current CH finds that its energy has fallen below  $E_{\text{sleep}}$ , it will first choose  $k$ , which is set to 3 in this paper, the nearest awake CMs  $[CM_1, CM_2, \dots, CM_k]$ , and calculate the corresponding energy states for each chosen CM. Then, the CM with the best energy state will be selected as the new CH. The result will be broadcast to all the CMs in the *Slot CHR*.

We believe that the CHR strategy can maintain the cluster's stability and keep the intercluster routing effectively since only the CMs near the current CH will be the candidates. On the other hand, by choosing the CM with the best energy state among the candidates, the CHR strategy can also better utilize the network energy. The CHR strategy may be executed several times in one working round. It can visibly improve the network performance.

## 5. Performance Evaluation

**5.1. Simulation Setting.** This section compares the performance of CPEH with three popular clustering protocols: LEACH, EAUCF, and MOFCA. We focus on a  $400\text{m} \times 400\text{m}$  sensing area with 400 energy harvesting sensors randomly deployed throughout the network. The network is assumed to be homogeneous, with a data aggregation ratio of 0.1.

To have an intuitive insight into the advantages of CPEH, we consider two different scenarios. Precisely, Scenario 1: the BS is located in the middle of the network, and the harvest rates of sensors range from  $25\ \mu\text{W}$  to  $250\ \mu\text{W}$  randomly. Scenario 2: the BS is located on the edge of the network, and the harvesting rates range from  $50\ \mu\text{W}$  to  $500\ \mu\text{W}$ . We show a typical deployment case in Figure 10.

We define each round as 0.1 h, and the network total simulation time is 72 h. To test the network performance under different network loads, we change the number of the frames in each round from 50 to 120. Ten different sensor deploy-

TABLE 2: The corresponding simulation parameters used in this paper.

Parameters	Values
Network area	[400 m, 400 m]
Number of sensors	400
Base station location	(200, 200), (200, 400)
Sensor initial energy	1 J
Sensor energy capacity	2 J
Number of frames	50-120
Sensor energy harvesting rate	( $25\ \mu\text{W}$ - $250\ \mu\text{W}$ ), ( $50\ \mu\text{W}$ , $500\ \mu\text{W}$ )
Data packet size	500 byte
Control packet size	25 byte
$E_{\text{elec}}$	50 nJ/bit
$E_{\text{fs}}$	10 pJ/bit/m <sup>2</sup>
$E_{\text{mp}}$	0.0013 pJ/bit/m <sup>4</sup>
$E_{\text{DA}}$	5 nJ/bit/signal
$d_0$	87 m
Competition radius	$R = 60\text{ m}$
Impact factor in fuzzy logic	$\alpha = 2, \beta = 1$
Number of ants	10
Number of iterations in ACO	20
Initial pheromone trail density	$\tau_{i,j}(0) = 10$
Impact factor in ACO	$\theta = 1, \nu = 3$
Constant factor $Q$	$Q = 0.05$
Pheromone decay coefficient	$\rho = 0.25$
Number of CHR candidates	$k = 3$

ments are randomly generated in each scenario to get the average result to eliminate the contingency. The relevant parameters are summarized in Table 2. The simulation is

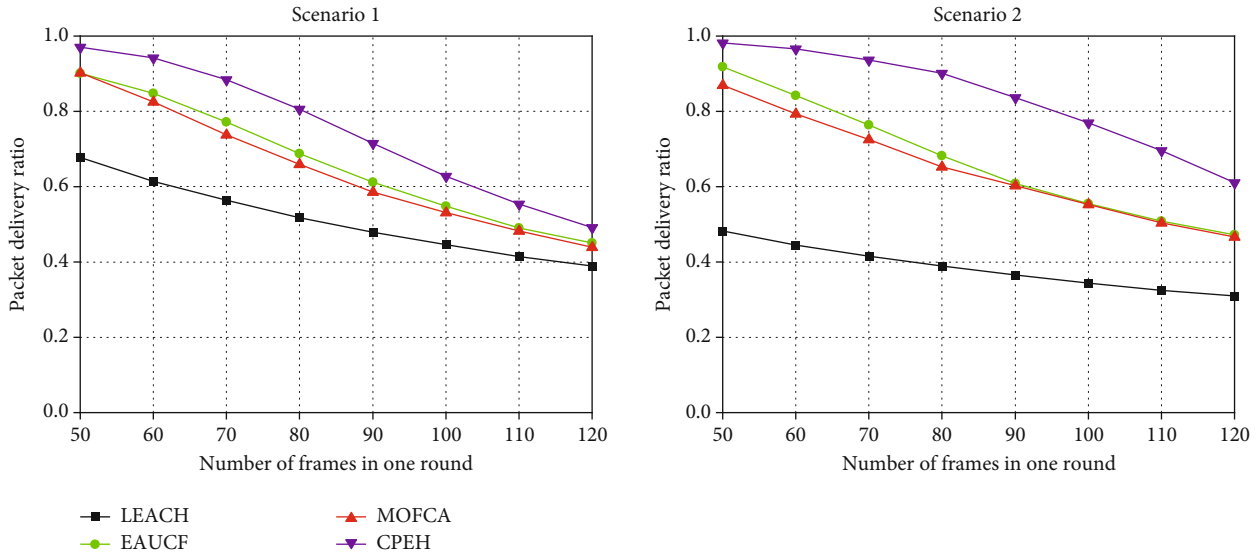


FIGURE 11: Network packer delivery ratio in different scenarios.

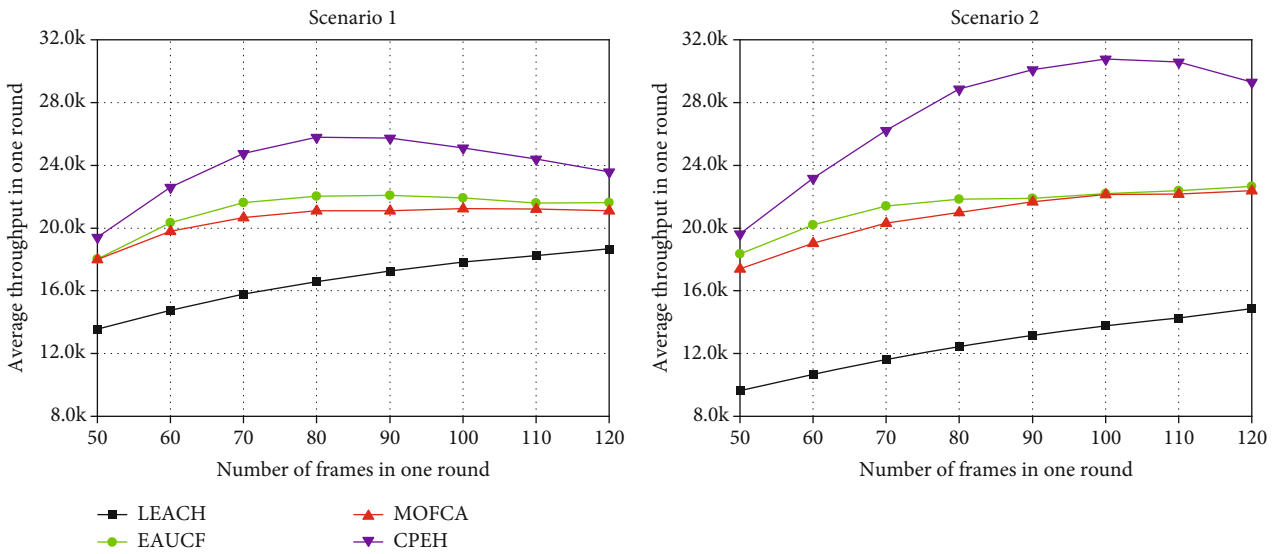


FIGURE 12: Average network throughput in one round in different scenarios.

executed by Matlab. All the energy dissipation of control packets and data packets is considered to guarantee veracity. The result is summarized in Section 5.2.

5.2. Simulation Result

5.2.1. Packet Delivery Ratio. Figure 11 shows the network packet delivery ratio under different frame numbers in two scenarios. With the increase of the frame number, the delivery ratio will gradually decrease under all the protocols, which can be easily explained. The more frame number, the more energy shall be needed to handle the data packets, leading to more packet losses owing to the energy shortage. However, it is noteworthy that CPEH outperforms others in all the conditions. For example, when the frame number is 80, in Scenario 1, CPEH can achieve a 55.6%, 17.1%, and 22.2%

increment over LEACH, EAUCF, and MOFCA separately. The improvement comes to 131.7%, 32%, and 38.2% in Scenario 2. We believe that the superiority of CPEH is owing to the more reasonable cluster and route construction.

5.2.2. Network Throughput. We illustrate the average network throughput in one round under different clustering protocols in Figure 12. When the frame number is low, increasing the frame number will improve the throughput under all the protocols. However, the throughput improvement becomes weak or even negative if the frame number is large. This phenomenon shows that we cannot improve the network throughput by naively assigning more frames in the working round. CPEH can achieve a 38.1%, 16.8%, and 21.5% and 107%, 35.9%, and 37.4% higher max

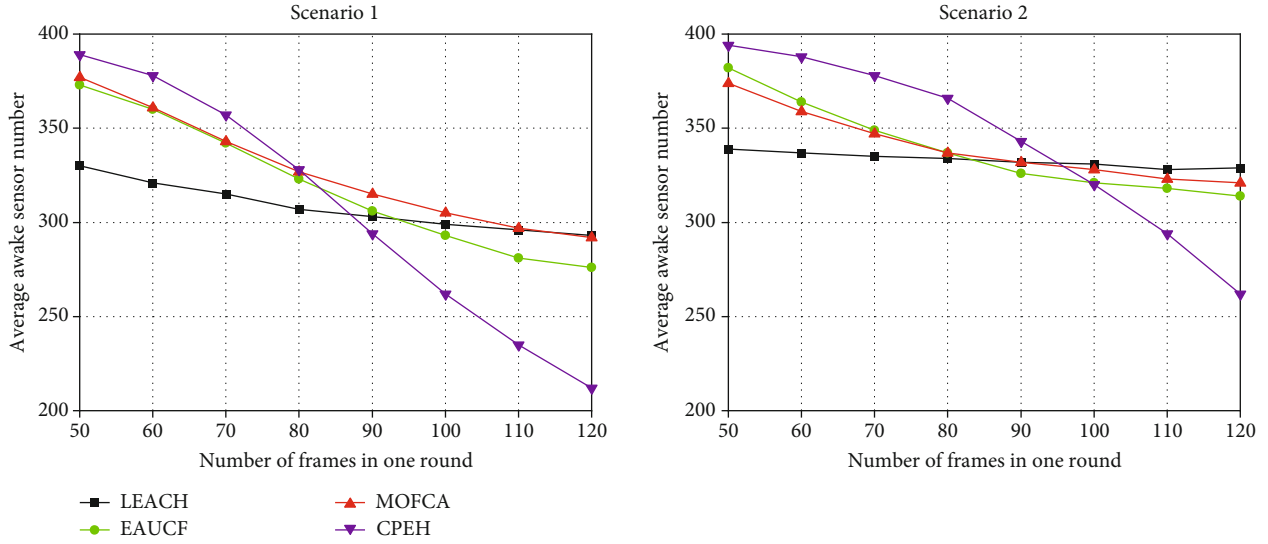


FIGURE 13: Average awake sensor number.

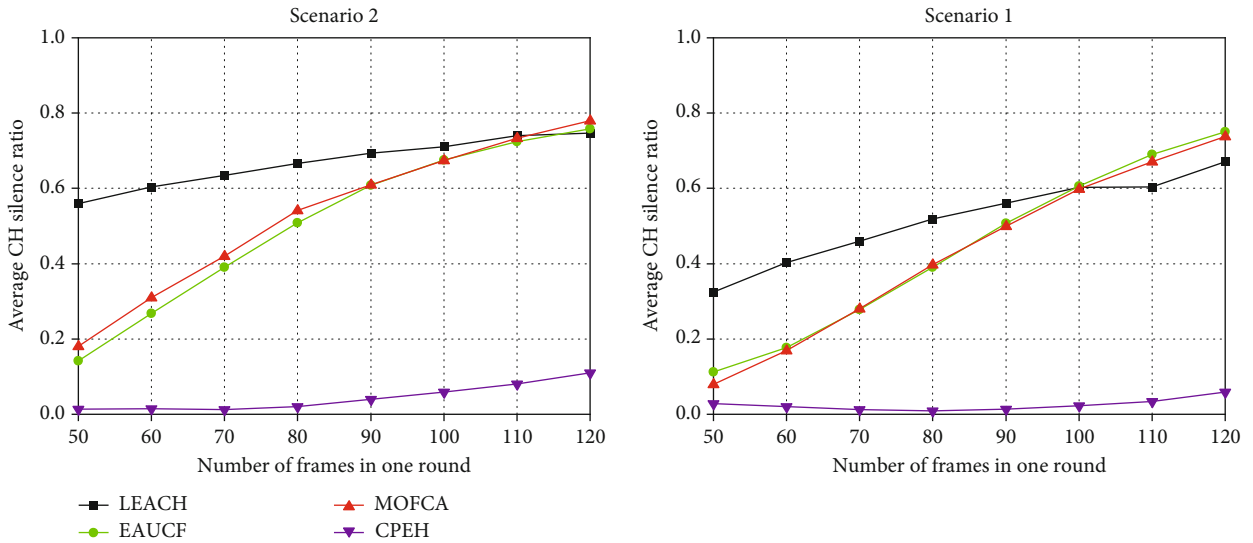


FIGURE 14: Average cluster dormancy ratio.

throughput in Scenario 1 and Scenario 2 than LEACH, EAUCF, and MOFCA, respectively, whereas the throughput and packet delivery ratio in CPEH has a trade-off relation, which shall be carefully considered.

**5.2.3. Awake Sensor Number and the CH Dormancy Ratio.** Since the sense is a cooperative work in the homogeneous network, the number of awake nodes will affect the sensing accuracy. We record the average number of awake nodes at every round beginning and summarize the result in Figure 13. The result shows that the average awake sensor number under CPEH will suffer a sharper decrease with the frame increase. Therefore, even though the CPEH can keep the network with more awake sensors when the frame number is relatively small, this index of CPEH becomes the worst under a large frame number. We think this phenomenon is

owing to the CHR strategy of CPEH. LEACH, EAUCF, and MOFCA do not consider the situation where the CH turns to sleep in the middle of the working round. In that case, even though CMs have enough energy to continue their work, they have to keep silent until the round ending, which will cause the next round to have more awake sensors. However, the CHR strategy can effectively solve this problem and let the awake CMs with enough energy continually work. We have to point out that even though CPEH may have fewer awake sensors in some conditions, it can still achieve a better delivery ratio and throughput.

Figure 14 shows the average CH dormancy ratio under different frame numbers. We can observe that the CHR strategy of CPEH can successfully avoid the cluster being silent in the working round. On the contrary, clusters under three other protocols may lose their function and become silent

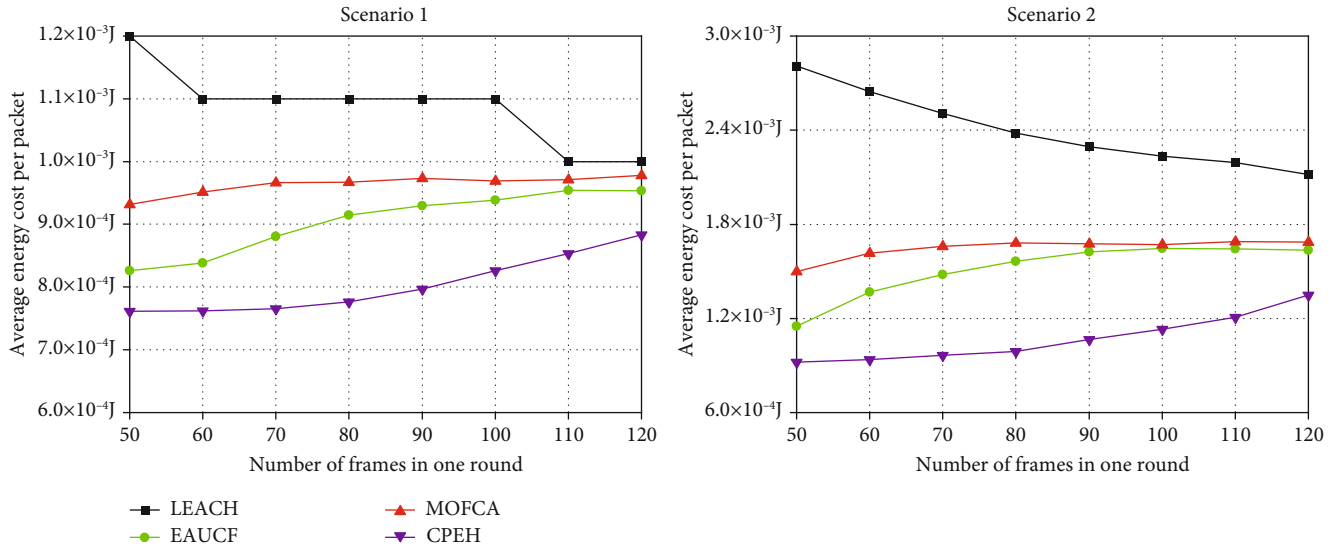


FIGURE 15: Average cluster silence ratio.

halfway, especially when the frame number is large. This conclusion can also prove that it is worthy of sacrificing few awake sensor numbers by taking the CHR strategy.

**5.2.4. Average Energy Cost per Packet.** We summarize the average energy cost of transmitting a data packet under different protocols and show the result in Figure 15. It is no doubt that CPEH is the most economical protocol in all network conditions, which proves its efficiency. We believe that, except for the more appropriate cluster construction and intercluster routing, the CHR strategy also contributes.

For EAUCF and MOFCA, when the CH falls into sleep during the working round, the routing topology will be changed. The last-hop and next-hop CHs of the sleeping CH will have to communicate directly, causing the long-range transmission and destroying the energy efficiency. Conversely, the CHR strategy can keep the intercluster routing topology robust and stable, leading to continuous, efficient intercluster communication.

## 6. Conclusion and Future Work

In this paper, we propose CPEH, which is a clustering protocol designed explicitly for EHWSN. CPEH mainly consists of two parts. The first part focuses on cluster construction. We adopt the fuzzy logic system to handle the uncertain nature of EHWSN and construct clusters more appropriately. The second part utilizes the ACO algorithm to optimize the intercluster routing, which can inherently achieve a better path than greedy algorithms used in most clustering protocols. We execute a comprehensive simulation of CPEH with some representative clustering protocols under different network conditions. The result proves that CPEH can always achieve the best performance in network delivery ratio and throughput. Furthermore, the CHR strategy of CPEH can effectively solve the cluster dormancy problem, ensuring the cluster works normally. The advantages of CPEH make it a suitable

protocol for Industry 4.0 era applications. For future work, we will extend CPEH to handle the multi-BS condition and consider sensors' movement.

## Data Availability

The authors do not list the raw simulation data in this paper owing to the spatial confined. Readers can get the detailed data by connecting to the corresponding author.

## Conflicts of Interest

The authors have declared that no competing interests exist concerning this study.

## Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under project contract nos. 61701082, 61701116, 61601093, 61971113, and 61901095, in part by National Key R&D Program under project contract nos. 2018YFB1802102 and 2018AAA0103203, in part by Guangdong Provincial Research and Development Plan in Key Areas under project contract nos. 2019B010141001 and 2019B010142001, in part by Sichuan Provincial Science and Technology Planning Program under project contract nos. 2018HH0034, 2019YFG0418, 2019YFG0120, 2020YFG0039, and 2018JY0246, in part by Ministry of Education China Mobile Fund Program under project contract no. MCM20180104, in part by Yibin Science and Technology Program—Key Projects under project contract nos. 2018ZSF001 and 2019GY001, and in part by Central University Business Fee Program under project contract no. A03019023801224 and Fundamental Research Funds for the Central Universities under Grant ZYGX2019Z022.

## References

- [1] S. Wang, J. Wan, D. Zhang, D. Li, and C. Zhang, "Towards smart factory for industry 4.0: a self-organized multi-agent system with big data based feedback and coordination," *Computer Networks*, vol. 101, pp. 158–168, 2016.
- [2] J. Wan, S. Tang, Z. Shu et al., "Software-defined industrial Internet of things in the context of industry 4.0," *IEEE Sensors Journal*, vol. 16, p. 1, 2016.
- [3] X. Li, D. Li, J. Wan, A. V. Vasilakos, C. F. Lai, and S. Wang, "A review of industrial wireless networks in the context of Industry 4.0," *Wireless Networks*, vol. 23, no. 1, pp. 23–41, 2017.
- [4] M. M. Krunz and T. Shu, "Coverage-time optimization for clustered wireless sensor networks," *IEEE/ACM Transactions on Networking (TON)*, vol. 18, no. 1, pp. 202–215, 2010.
- [5] M. Youssef, A. Youssef, and M. Younis, "Overlapping multi-hop clustering for wireless sensor networks," *IEEE Transactions on Parallel & Distributed Systems*, vol. 20, no. 12, pp. 1844–1856, 2009.
- [6] P. Ding, J. A. Holliday, and A. Celik, "Distributed energy-efficient hierarchical clustering for wireless sensor networks," in *IEEE International Conference on Distributed Computing in Sensor Systems*, Berlin, Heidelberg, 2005.
- [7] N. Sabor, M. Abo-Zahhad, S. Sasaki, and S. M. Ahmed, "An unequal multi-hop balanced immune clustering protocol for wireless sensor networks," *Applied Soft Computing*, vol. 43, pp. 372–389, 2016.
- [8] Z. Hanzalek and P. Jurcik, "Energy efficient scheduling for cluster-tree wireless sensor networks with time-bounded data flows: application to IEEE 802.15.4/ZigBee," *IEEE transactions on industrial informatics*, vol. 6, no. 3, pp. 438–450, 2010.
- [9] M. L. Ku, W. Li, Y. Chen, and K. R. Liu, "Advances in energy harvesting communications: past, present, and future challenges," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 1384–1412, 2017.
- [10] S. Ulukus, A. Yener, E. Erkip et al., "Energy harvesting wireless communications: a review of recent advances," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 3, pp. 360–381, 2015.
- [11] S. Sudevalayam and P. Kulkarni, "Energy harvesting sensor nodes: survey and implications," *IEEE Communications Surveys & Tutorials*, vol. 13, no. 3, pp. 443–461, 2011.
- [12] S. Peng and C. P. Low, "Prediction free energy neutral power management for energy harvesting wireless sensor nodes," *Ad Hoc Networks*, vol. 13, pp. 351–367, 2014.
- [13] D. Hasenfratz, A. Meier, C. Moser, J. J. Chen, and L. Thiele, "Analysis, comparison, and optimization of routing protocols for energy harvesting wireless sensor networks," in *IEEE International Conference on Sensor Networks*, Newport Beach, CA, USA, 2010.
- [14] B. Baranidharan and B. Santhi, "DUCF: distributed load balancing unequal clustering in wireless sensor networks using fuzzy approach," *Applied Soft Computing*, vol. 40, pp. 495–506, 2016.
- [15] J. M. Kim, S. H. Park, Y. J. Han, and T. M. Chung, "CHEF: cluster head election mechanism using fuzzy logic in wireless sensor networks," in *2008 10th International Conference on Advanced Communication Technology*, vol. 1, pp. 654–659, Gangwon, Korea (South), 2008.
- [16] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient protocol for wireless microsensor networks," in *Proceedings of the 33rd annual Hawaii international conference on system sciences*, Maui, HI, USA, 2000.
- [17] M. Ye, C. Li, G. Chen, and J. Wu, "EECS: an energy efficient clustering scheme in wireless sensor networks, conference IEEE international performance," in *PCCC 2005. 24th IEEE International Performance, Computing, and Communications Conference*, Phoenix, AZ, USA, 2005.
- [18] O. Younis and S. Fahmy, "HEED: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks," *IEEE Transactions on Mobile Computing*, vol. 3, no. 4, pp. 366–379, 2004.
- [19] G. Chen, C. Li, M. Ye, and J. Wu, "An unequal cluster-based routing protocol in wireless sensor networks," *Wireless Networks*, vol. 15, no. 2, pp. 193–207, 2009.
- [20] N. M. A. Latiff, C. C. Tsimenidis, and B. S. Sharif, "Energy-aware clustering for wireless sensor networks using particle swarm optimization," in *2007 IEEE 18th international symposium on personal, indoor and mobile radio communications*, pp. 1–5, Athens, Greece, 2007.
- [21] Y. Han, G. Li, R. Xu, J. Su, J. Li, and G. Wen, "Clustering the wireless sensor networks: a meta-heuristic approach," *IEEE Access*, vol. 8, pp. 214551–214564, 2020.
- [22] H. Bagci and A. Yazici, "An energy aware fuzzy approach to unequal clustering in wireless sensor networks," *Applied Soft Computing*, vol. 13, no. 4, pp. 1741–1749, 2013.
- [23] S. A. Sert, H. Bagci, and A. Yazici, "MOFCA: multi-objective fuzzy clustering algorithm for wireless sensor networks," *Applied Soft Computing Journal*, vol. 30, pp. 151–165, 2015.
- [24] S. Peng, T. Wang, and C. P. Low, "Energy neutral clustering for energy harvesting wireless sensors networks," *Ad Hoc Networks*, vol. 28, pp. 1–16, 2015.
- [25] J. Li and D. Liu, "DPSO-based clustering routing algorithm for energy harvesting wireless sensor networks," in *International Conference on Wireless Communications & Signal Processing*, pp. 1–5, Nanjing, China, 2015.
- [26] Y. Dong, J. Wang, B. Shim, and D. I. Kim, "DEARER: a distance and-energy-aware routing with energy reservation for energy harvesting wireless sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 12, pp. 3798–3813, 2016.
- [27] T. Ross, *Fuzzy-Logik. Fuzzy Logic with Engineering Applications*, McGraw-Hill, Inc, 1995.
- [28] M. Dorigo and C. Blum, "Ant colony optimization theory: a survey," *Theoretical Computer Science*, vol. 344, no. 2–3, pp. 243–278, 2005.
- [29] J. Bai, G. K. Yang, Y. W. Chen, L. S. Hu, and C. C. Pan, "A model induced max-min ant colony optimization for asymmetric traveling salesman problem," *Applied Soft Computing*, vol. 13, no. 3, pp. 1365–1375, 2013.

## Research Article

# Coword and Cluster Analysis for the Romance of the Three Kingdoms

Chao Fan<sup>1,2</sup> and Yu Li<sup>1,2</sup>

<sup>1</sup>The School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi 214122, China

<sup>2</sup>Jiangsu Key Laboratory of Media Design and Software Technology, Jiangnan University, Wuxi 214122, China

Correspondence should be addressed to Chao Fan; fanchao@jiangnan.edu.cn

Received 1 March 2021; Revised 12 March 2021; Accepted 19 March 2021; Published 1 April 2021

Academic Editor: Shan Zhong

Copyright © 2021 Chao Fan and Yu Li. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The *Romance of the Three Kingdoms* (RTK) is a classical Chinese historical novel by Luo Guanzhong. This paper establishes a research framework of analyzing the novel by utilizing coword and cluster analysis technology. At the beginning, we segment the full text of the novel, extracting the names of historical figures in the RTK novel. Based on the coword analysis, a social network of historical figures is constructed. We calculate several network features and enforce the cluster analysis. In addition, a modified clustering method using edge betweenness is proposed to improve the effect of clustering. Finally, both quantified and visualized results are displayed to confirm our approach.

## 1. Introduction

The *Romance of the Three Kingdoms*, written by Luo Guanzhong, is generally considered to be one of the four great classical novels in Chinese literature. It describes the turbulent years from the end of the Han dynasty to the Three Kingdoms (Wei, Shu, and Wu) era in Chinese history. More than 1000 personalities are vividly portrayed in the historical novel.

In this research, text of original novel is divided into a number of sentences. According to coword analysis, there is a certain intrinsic relationship between the two words when they appear in the same document. Thus, we calculated the frequency of cooccurrences for two names in a sentence. The character name is reckoned as the node and the cooccurrence as the link, so that an undirected network can be established. Furthermore, various network features are computed to analyze relationships of characters in the novel. Cluster analysis is employed to explore the hierarchical structure of RTK. Finally, an improved clustering algorithm by cutting high-betweenness edges is proposed, which performs better than the common approach in clustering effect.

This manuscript is organized as follows. Section 2 gives related work of this paper. Data preparation is discussed in Section 3. Sections 4 and 5 express the network feature anal-

ysis, cluster analysis, experiments, and the analysis of results. Conclusions are drawn in Section 6.

## 2. Related Work

Early research about the RTK concentrates on qualitative analysis, such as the writing style, genealogy, and characters. Later, a quantitative approach was adopted to analyze the novel. Coword analysis is such a method of importance, which was first devised by French scholars and introduced into the information science field by Callon [1]. According to the theory of coword analysis, there is a close connection between two words when they appear in a sentence. More cooccurrences of the two words indicate the closer relationship between them. In this paper, we consider the cooccurrence of character names in a sentence of the RTK novel.

Numerous researches on literature analysis have been done based on the technologies of coword analysis. Ravikumar et al. [2] inspect 959 articles in scientometrics based on the coword analysis approach and find that the topics in publication are changing to new themes. As for the medical literature, there is a study utilizing this tool to process them over a span of thirty years [3]. Another work focuses on past themes and future trends in medical



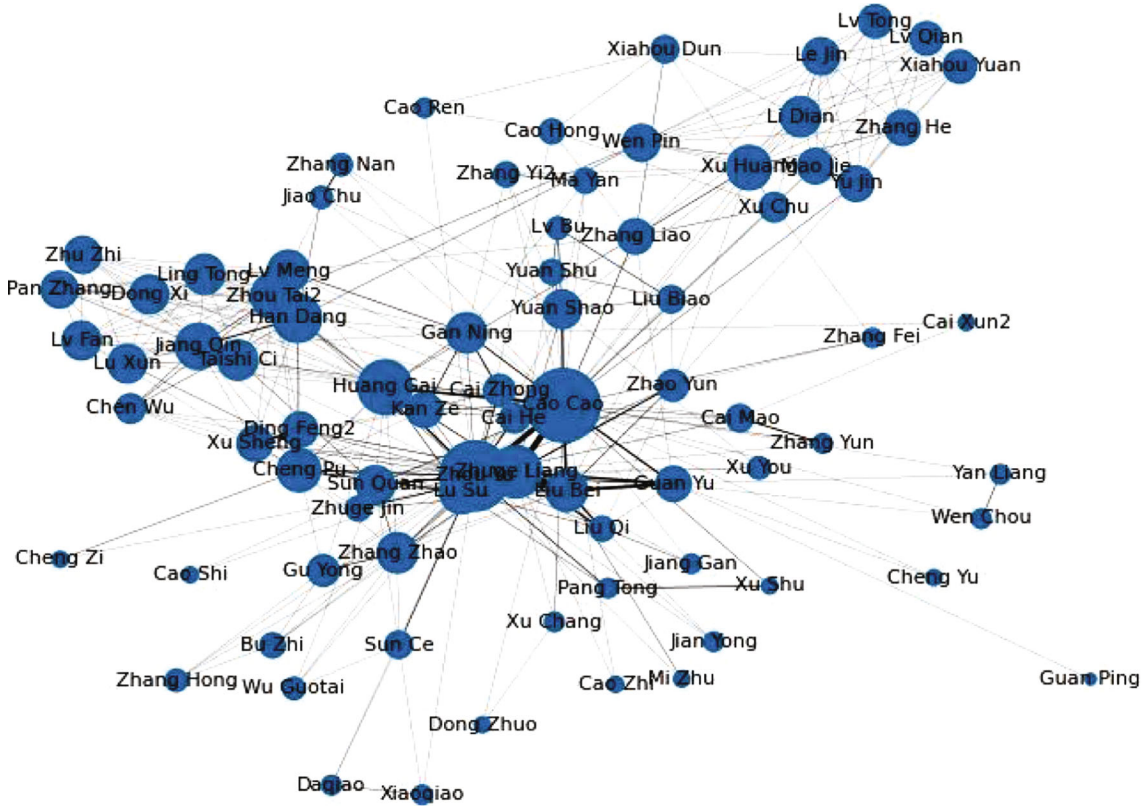


FIGURE 1: A network of character names (top 80 in node frequency).

tourism research [4]. Employing the cword analysis, some researchers attempt to identify the themes and trends of main knowledge areas including engineering, health, public administration, and management [5]. Moreover, a cword network is established to analyze the relationship of characters in the *Dream of the Red Chamber* [6]. Wang et al. build a similar network for the *Romance of the Three Kingdoms* [7].

After creating a social network based on cword analysis, the cluster analysis is carried out by performing a hierarchical clustering algorithm. Two types of algorithm are often implemented when moving up the hierarchy. The divisive approach of clustering reckons all data as one cluster and performs splits, which is used in many research [8]. Nevertheless, the agglomerative hierarchical clustering is a bottom-up method with many variants [9]. It merges the two most similar clusters at each time. The agglomerative method is exploited in this work because it can provide a visual expression of the clustering results.

### 3. Data Preparation

**3.1. Building RTK Corpus and Preprocessing.** As many data of the novel can be downloaded from the Internet, we selected a high-quality text document (<https://72k.us/file/22215238-408791478>) in Chinese character, establishing the RTK corpus by cleaning the original data. Some words with errors were modified, and the wrong punctuations were removed manually.

The raw text is preprocessed using the natural language processing toolkit ICTCLAS (<http://ictclas.nlpir.org/>). We

acquired a name list of RTK characters through the Internet and added it to the dictionary of ICTCLAS. Then, the lexical analysis is executed to segment Chinese sentences into words where names of characters can be found.

**3.2. Creation of Character Name Network.** Based on cword analysis, an undirected network of character names can be created by counting the cooccurrences of two names in sentences. We treated full name, its courtesy name, and abbreviated name as one name. For example, “Cao Cao” is equal to “Cao Mengde” and “Mengde,” which means the three names refer to a single person of “Cao Cao.”

The final constructed network of character names has 1,133 nodes and 5,844 links. As depicted in Figure 1, the size of a node indicates the count of the character name in the novel and the thickness of a link corresponds to the frequency of two characters that appear together.

## 4. Network Feature Analysis

**4.1. Degree Distribution.** As the degree of a node is the number of links adjacent to it, the degree distribution is the probability distribution of these degrees. A power index  $\gamma$  can be used to describe the curve if the network’s degree distribution follows a power-law distribution.

For the network of RTK characters, the top ten characters of the highest degree are Cao Cao, Liu Bei, Zhuge Liang, Sun Quan, Zhao Yun, Guan Yu, Yuan Shao, Sima Yi, Lv Bu, and Wei Yan. The average degree of the network is 10.31, and the degree distribution can be illustrated in Figure 2. It emerges

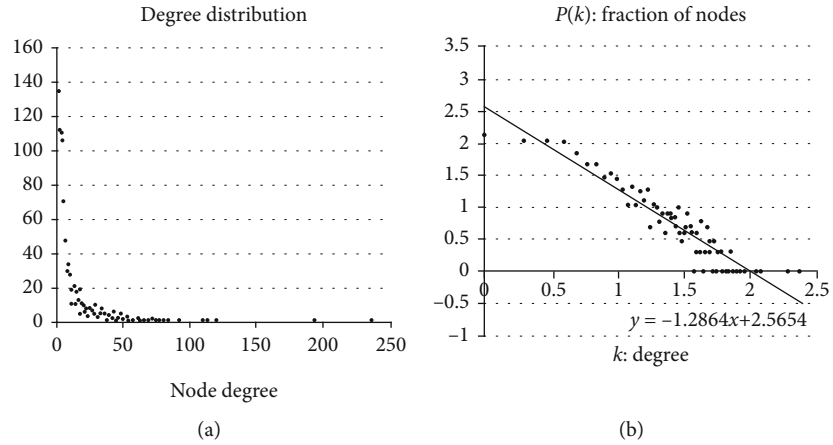


FIGURE 2: Degree distribution and power-law degree distribution on a log-log scale.

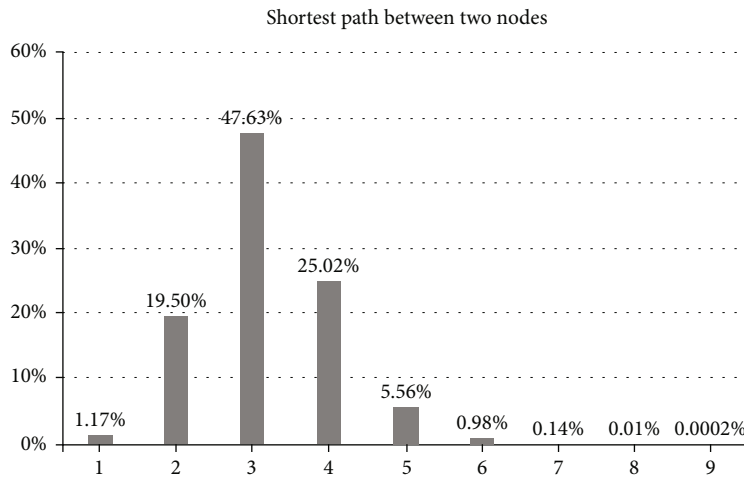


FIGURE 3: Distribution of shortest-path length.

to be a heavy-tailed distribution (see Figure 2(a)). As the data can be approximated with a linear function  $y = -1.2864x + 2.5654$  on a log-log scale in Figure 2(b), we conclude that the degree distribution follows a power-law distribution.

**4.2. Average Shortest-Path Length.** The shortest path between two nodes is a path where the number of links is minimized. Accordingly, the length of the shortest path is the number of links that the path contains. A sum of all shortest-path length divided by the number of links is the average shortest-path length.

The average shortest-path length of the RTK network is 3.1743. Hence, one character can be connected to others in three steps on average, which means any two characters are “three-degree separation.”

The distance of the largest shortest path in the network is called diameter. In this paper, the RTK network’s diameter is 9. One path of the diameter is from Liu Ai to Zhang Shang: Liu Ai, Wang Li, Dong Zhao, Cao Hong, Cao Cao, Sima Yan, Yang Hu, Du Yu, Lu Jing, and Zhang Shang. The distribution of the shortest-path length between any two characters can be illuminated in Figure 3. According to the figure,

47.63% of the shortest-path length in the RTK network is 3 and about 92.15% is between length 2 and length 4.

**4.3. Clustering Coefficient.** A clustering coefficient [10, 11] measures the extent to which a network’s nodes tend to cluster together. The clustering coefficient of node  $x$  can be given by

$$C_x = \frac{2E_x}{k_x(k_x - 1)}. \quad (1)$$

$E_x$  is the existing links among neighbors of node  $x$ . As  $k_x$  is a degree of node  $x$ ,  $(1/2)k_x(k_x - 1)$  represents the number of potential links for node  $x$ ’s neighbors. Therefore, the average value for all  $C_x$  is the clustering coefficient of the whole network.

$$C = \frac{1}{N} \sum_x C_x. \quad (2)$$

A random network is produced by an Erdős-Rényi (ER) model utilizing the same number of nodes and links as the RTK network. The comparison between random network

TABLE 1: Comparison between RTK and random network.

	Number of nodes	Number of links	Average degree	Average shortest-path length	Clustering coefficient
RTK network	1,133	5,844	10.3159	3.1743	0.5306
Random network	1,133	5,844	10.3159	3.2702	0.0082

TABLE 2: Comparison of three subnetworks and the whole network.

	Density	Clustering coefficient	Average shortest-path length	Diameter
Shu	0.1652	0.6635	2.0563	4
Wu	0.1099	0.5845	2.3054	5
Wei	0.0803	0.6217	2.5953	6
The whole network	0.0091	0.5306	3.1743	9

TABLE 3: Top 10 characters in rank with the highest centrality.

Ranking	Degree centrality	Betweenness centrality	Closeness centrality
1	Cao Cao (0.2094)	Cao Cao (0.1751)	Cao Cao (0.4528)
2	Liu Bei (0.2085)	Liu Bei (0.1304)	Liu Bei (0.4442)
3	Zhuge Liang (0.1714)	Zhuge Liang (0.1093)	Zhuge Liang (0.4313)
4	Sun Quan (0.1060)	Sun Quan (0.0695)	Sun Quan (0.4073)
5	Zhao Yun (0.0998)	Sima Yi (0.0430)	Guan Yu (0.3969)
6	Guan Yu (0.0972)	Zhao Yun (0.0413)	Zhao Yun (0.3963)
7	Yuan Shao (0.0813)	Liu Shan (0.0402)	Sima Yi (0.3924)
8	Sima Yi (0.0742)	Guan Yu (0.0375)	Wei Yan (0.3856)
9	Lv Bu (0.0716)	Yuan Shao (0.0369)	Yuan Shao (0.3842)
10	Wei Yan (0.0707)	Jiang Wei (0.0357)	Cao Ren (0.3824)

and RTK network is shown in Table 1. The RTK network is a small-world network because it has a larger clustering coefficient as well as a smaller average shortest-path length compared with a random network.

We choose the characters who clearly belong to the three groups of Wei, Shu, and Wu and calculate the network features of the three kingdoms, respectively. The results are summarized in Table 2.

The character relationship networks within three groups have high clustering coefficients and small average shortest-path lengths. Consequently, all of the three subnetworks are “small-world” networks. From the Shu to Wu and Wei, the density and clustering coefficient of the subnetworks decrease sequentially except for the clustering coefficient of Wu. On the contrary, the average shortest-path length and diameter increase successively. This reflects a decrease in the closeness of the connections among the groups. In other words, the connections among characters in Wei are less closely than Wu and Shu.

**4.4. Density.** The density of a network shows the ratio of links, which can be simply calculated by formula (3).  $N$  and  $E$  are the number of nodes and links. It describes the portion of all possible links in a network that are actual connections.

The value is a fraction between 0 and 1. As the density of the RTK network is 0.0091, it is a sparse network.

$$d = \frac{2E}{N(N-1)}. \quad (3)$$

**4.5. Centrality.** The centrality measures the importance of nodes, containing degree centrality, betweenness centrality, and closeness centrality.

Degree centrality is a measure of centrality based on degree. A high-degree node is a local center within the network. Betweenness centrality expresses the extent that the node falls on the shortest path between other pairs of nodes. A node with a high betweenness is capable of controlling the interactions between two nonadjacent nodes [5]. Closeness centrality is a measure of the average shortest distance from each node to each other node. It evaluates the closeness that a node is to all the other nodes [3].

Three centralities of characters in the RTK network are calculated, respectively. Table 3 gives the top ten characters of the highest centrality. The value of centrality is listed in parentheses. From Table 3, we can find eight names listed in three centralities: Cao Cao, Liu Bei, Zhuge Liang, Sun

TABLE 4: Cooccurrence matrix of main characters.

Cooccurrence	Liu Bei	Cao Cao	Sun Quan	Zhuge Liang	Guan Yu	Zhang Fei
Liu Bei	541	112	58	190	106	75
Cao Cao	112	275	39	50	58	16
Sun Quan	58	39	145	28	18	2
Zhuge Liang	190	50	28	336	43	25
Guan Yu	106	58	18	43	272	47
Zhang Fei	75	16	2	25	47	165

TABLE 5: Ochiai similarity matrix of main characters.

Cooccurrence	Liu Bei	Cao Cao	Sun Quan	Zhuge Liang	Guan Yu	Zhang Fei
Liu Bei	1	0.290371	0.207083	0.445641	0.276327	0.251027
Cao Cao	0.290371	1	0.195305	0.164488	0.212069	0.075112
Sun Quan	0.207083	0.195305	1	0.126854	0.090637	0.01293
Zhuge Liang	0.445641	0.164488	0.126854	1	0.142238	0.106176
Guan Yu	0.276327	0.212069	0.090637	0.142238	1	0.221856
Zhang Fei	0.251027	0.075112	0.01293	0.106176	0.221856	1

TABLE 6: The clustering result of the RTK network ( $k$  is the final number of hierarchical clusters).

$k$	Precision	Recall	$F$ score
...	...	...	...
11	43.83%	78.90%	56.35%
12	47.08%	78.90%	58.97%
13	71.10%	75.00%	73.00%
14	71.10%	75.00%	73.00%
<b>15</b>	<b>87.66%</b>	<b>73.38%</b>	<b>79.89%</b>
16	87.66%	62.99%	73.30%
17	87.66%	59.09%	70.60%
18	87.66%	50.97%	64.46%
...	...	...	...

Quan, Zhao Yun, Guan Yu, Yuan Shao, and Sima Yi. They are in a significant position in the character network.

## 5. Cluster Analysis

*5.1. Cooccurrence and Similarity Matrix.* The cooccurrence matrix measures the frequency that two characters appear together. A cooccurrence matrix of main characters in the RTK network is presented in Table 4. It is a symmetric matrix, and data on the diagonal show the frequencies of characters that appear in text.

The cooccurrence of two characters cannot be used as the similarity because it is greatly affected by frequency. We normalize the cooccurrence matrix utilizing the Ochiai coefficient [12] and obtain the similarity matrix. Ochiai coefficient is defined by

$$K = \frac{n(A \cap B)}{\sqrt{n(A) \times n(B)}}. \quad (4)$$

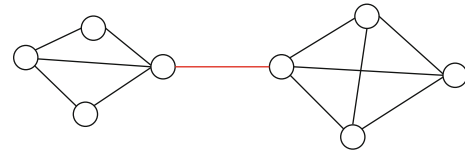


FIGURE 4: A link with a high edge betweenness.

TABLE 7: The clustering result of the RTK network ( $k$  is the final number of hierarchical clusters).

Number of removals	Precision	Recall	$F$ score
0	87.66%	73.38%	79.89%
5	87.66%	73.38%	79.89%
10	88.31%	73.38%	80.15%
15	88.64%	74.35%	80.87%
20	88.64%	74.35%	80.87%
25	88.64%	74.35%	80.87%
30	88.64%	74.35%	80.87%
35	88.64%	74.35%	80.87%
40	88.96%	73.70%	80.62%
45	88.96%	73.38%	80.42%
50	89.94%	73.05%	80.62%
55	89.94%	72.08%	80.02%
60	47.73%	90.58%	62.52%
...	...	...	...

As  $A$  and  $B$  are sets,  $n(A)$  is the number of elements in  $A$  and  $n(A \cap B)$  is the number of cooccurrence. The similarity matrix calculated by the Ochiai coefficient is described in Table 5.

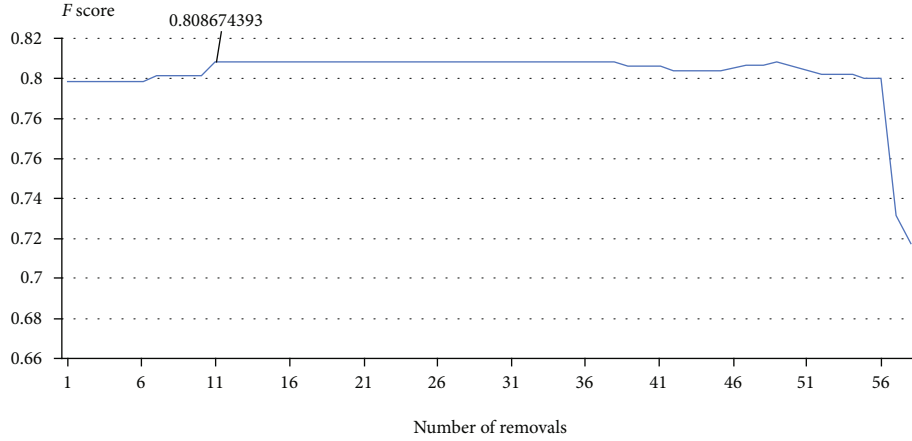


FIGURE 5: The change of  $F$  score according to the number of removals.

## 5.2. Hierarchical Clustering

**5.2.1. Clustering Algorithm.** An agglomerative hierarchical clustering algorithm utilizing the Ochiai similarity matrix is implemented to complete the task of cluster analysis. It is a bottom-up approach. Initially, each node is treated as a single cluster. Two clusters with the largest Ochiai similarity are combined into a new bigger cluster. The clustering algorithm stops when it achieves a setting threshold or there is only one cluster left. The similarity between two clusters is defined as the average similarity between each of their nodes.

**5.2.2. Evaluation.** The P-IP scores [13] are adopted to measure the clustering result. There are  $m$  character names and  $n$  clusters. Suppose  $C_{ij}$  is the number of character names marked with label  $j$  for character name  $i$ , where  $j = \arg \max_k \{C_{ik}\}$ . The precision and recall of character name  $i$  can be given by

$$P_i = \frac{C_{ij}}{\sum_{l=1}^m C_{lj}}, \quad (5)$$

$$R_i = \frac{C_{ij}}{\sum_{k=1}^n C_{ik}}.$$

Thus, the  $F$  score is calculated by

$$F_i = \frac{2P_i R_i}{P_i + R_i}. \quad (6)$$

The overall precision, recall, and  $F$  score are the averages of corresponding values. Moreover, the gold standard is built by marking the character name with a specific kingdom tag. For example, Cao Cao is tagged with “Wei” and Liu Bei is tagged with “Shu.” Finally, 308 character names with definite kingdom tags are secured for cluster analysis.

**5.2.3. Clustering Result.** The result of hierarchical clustering is illustrated in Table 6. The  $F$  score achieves the best value of 79.89% when the number of clusters  $k$  is 15.

**5.3. Improved Clustering Algorithm.** In the RTK network, some characters play a vital role in interconnections of different kingdoms, like “Lu Su” between Wu and Shu, “Huang Gai” between Wu and Wei. These characters have a high betweenness according to the definition of betweenness (see Section 4.5). Further, the node betweenness can be extended to “edge betweenness” [14]. The link with a high edge betweenness is often a bridge between different clusters (see red link in Figure 4). Therefore, removing these high-betweenness links by setting a similarity of 0 will reduce the intercluster similarity and improve the clustering result eventually. The removal operation can be introduced as preprocessing before conducting the cluster analysis.

The improved clustering algorithm using edge betweenness is executed, and the result is displayed in Table 7. When the number of removals is zero, it is the baseline of the original algorithm. With an adequate removing operation, the  $F$  score reaches a peak of 80.87%. Nevertheless, removing too many links will destroy the whole network and make the  $F$  score decline dramatically (see Figure 5).

**5.4. Analysis.** Data visualization is also given to display the characteristics of historical figures in the RTK network. As hierarchical clustering can be depicted as a tree-based visual dendrogram, we visualize the character relationship in the RTK novel from Chapter 43 to 50, which is a period describing “the battle of Red Cliffs” (see Figure 6).

As can be seen from Figure 6, six parts can be divided manually. H1 and H3 are groups containing characters from “Wu,” like Sun Quan and Sun Ce. H2 encompasses main characters from “Shu” and “Wu” in the battle of Red Cliffs: Liu Bei, Guan Yu, Zhuge Liang, Zhou Yu, Lu Su, etc. However, there are two exceptions: Cao Cao and Cheng Yu, because they are highly connected with other main characters in the battle of Red Cliffs. Further, H1, H3, and H2 merge into a bigger cluster in the hierarchical clustering because these characters are from the alliance of “Wu” and “Shu” against Cao’s army.

On the other hand, H5 is composed of characters from a large group “Wei,” including Xiahou Dun, Xiahou Yuan, Cao Ren, and Cao Hong. H6 includes few characters from “Shu”

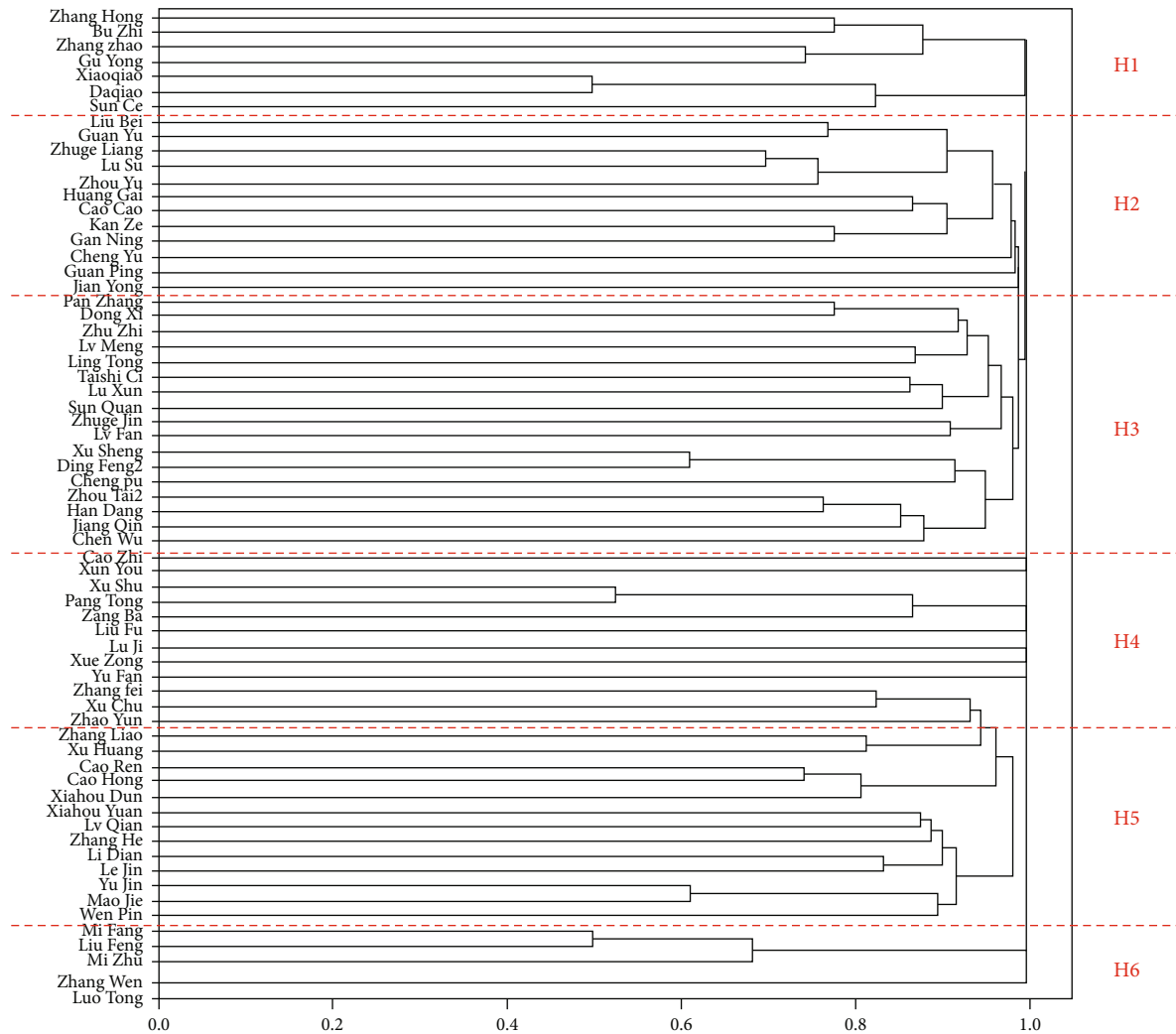


FIGURE 6: Dendrogram of clustering result for the period of “the battle of Red Cliffs.”

or “Wu.” H4 is not a cluster, and it contains a number of characters from different kingdoms.

**6. Conclusions**

This paper developed a general framework for analyzing the character relationship in the novel. The *Romance of the Three Kingdoms* is taken as the object of analysis. At first, the raw text of the RTK novel is processed with NLP tools and character names are recognized by lexical analysis. Then, a character name network is created based on coword analysis. After building the network, several network features are calculated such as degree distribution, average shortest-path length, and clustering coefficient. Besides, cluster analysis is conducted and it helps to better understanding of the hierarchical structure for characters in the RTK novel. A modified clustering algorithm using edge betweenness is proposed to improve the effect of clustering. Finally, visualization of results is completed to analyze the hierarchical clustering.

There are some limitations of the proposed method since coword analysis does not necessarily reflect the true meaning of character relationship. However, our approach can study the main characters quantitatively and comprehend character relationship from another perspective. Hence, it is a valuable research direction.

Subsequent work will study the meaning of pronouns because they represent different characters in different situations. Further, place names and institutions will be taken into consideration in the future.

**Data Availability**

The original dataset used in this work is available from the corresponding author on request.

**Conflicts of Interest**

The authors declare no conflicts of interest.

## Acknowledgments

This work was supported by the Youth Foundation of Basic Science Research Program of Jiangnan University, 2019 (No. JUSRP11962), and the High-Level Innovation and Entrepreneurship Talents Introduction Program of Jiangsu Province of China, 2019.

## References

- [1] Q. Zhu, X. Peng, and X. Liu, "Research topics in social computing area based on co-word analysis," *Information Studies: Theory & Application*, vol. 12, pp. 7–11, 2012.
- [2] S. Ravikumar, A. Agrahari, and S. N. Singh, "Mapping the intellectual structure of scientometrics: a co-word analysis of the journal *Scientometrics* (2005-2010)," *Scientometrics*, vol. 102, no. 1, pp. 929–955, 2015.
- [3] D. Nguyen, "Mapping knowledge domains of non-biomedical modalities: a large-scale co-word analysis of literature 1987-2017," *Social Science & Medicine*, vol. 233, pp. 1–12, 2019.
- [4] A. de la Hoz-Correa, F. Muñoz-Leiva, and M. Bakucz, "Past themes and future trends in medical tourism research: a co-word analysis," *Tourism Management*, vol. 65, pp. 200–211, 2018.
- [5] D. Corrales-Garay, M. Ortiz-de-Urbina-Criado, and E. M. Mora-Valentín, "Knowledge areas, themes and future research on open data: a co-word analysis," *Government Information Quarterly*, vol. 36, no. 1, pp. 77–87, 2019.
- [6] C. Fan, "Research on relationships of characters in the dream of the red chamber based on co-word analysis," *ICIC Express Letters Part B: Applications*, vol. 11, no. 5, pp. 1–8, 2020.
- [7] Y. Wang, J. Yu, and C. Zhao, "Research on application of co-word analysis on relationships of characters in the romance of the three kingdoms," *Information Research*, vol. 7, pp. 52–56, 2017.
- [8] A. Ishizaka, B. Lokman, and M. Tasiou, "A stochastic multi-criteria divisive hierarchical clustering algorithm," *Omega*, vol. 11, 2020.
- [9] N. Liu, Z. Xu, X. J. Zeng, and P. Ren, "An agglomerative hierarchical clustering algorithm for linear ordinal rankings," *Information Sciences*, vol. 557, pp. 170–193, 2021.
- [10] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.
- [11] C. Fan and F. Toriumi, "High-modularity network generation model based on the multilayer network," *Transactions of the Japanese Society for Artificial Intelligence*, vol. 32, no. 6, pp. B-H42\_1–B-H4211, 2017.
- [12] Q. Zhou and L. Leydesdorff, "The normalization of occurrence and co-occurrence matrices in bibliometrics using cosine similarities and Ochiai coefficients," *Journal of the Association for Information Science & Technology*, vol. 67, no. 11, pp. 1–25, 2016.
- [13] A. Hotho, S. Staab, and G. Stumme, "WordNet improves text document clustering," *Proceedings of the SIGIR 2003 Semantic Web Workshop*, pp. 541–544, 2003.
- [14] M. Givan and M. E. J. Newman, "Community structure in social and biological networks," *Proceedings of the national academy of sciences*, vol. 99, no. 12, pp. 7821–7826, 2002.

## Research Article

# Microgrid Group Control Method Based on Deep Learning under Cloud Edge Collaboration

Yazhe Mao,<sup>1</sup> Baina He ,<sup>1</sup> Deshun Wang,<sup>2</sup> Renzhuo Jiang,<sup>1</sup> Yuyang Zhou,<sup>1</sup> Xingmin He,<sup>1</sup> Jingru Zhang,<sup>1</sup> and Yanchen Dong<sup>1</sup>

<sup>1</sup>College of Electric and Electronic Engineering, Shandong University of Technology, Zibo 255000, China

<sup>2</sup>China Electric Power Research Institute, Nanjing 210003, China

Correspondence should be addressed to Baina He; hebaina@sdut.edu.cn

Received 2 December 2020; Revised 3 February 2021; Accepted 5 March 2021; Published 18 March 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Yazhe Mao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Aiming at the economic benefits, load fluctuations, and carbon emissions of the microgrid (MG) group control, a method for controlling the MG group of power distribution Internet of Things (IoT) based on deep learning is proposed. Firstly, based on the cloud edge collaborative power distribution IoT architecture, combined with distributed generation, electric vehicles (EV), and load characteristics, the MG system model in the power distribution IoT is established. Then, a deep learning algorithm is used to train the features of the data model on the edge side. Finally, the group control strategy is adopted in the power distribution cloud platform to reasonably regulate the coordinated output of multiple energy sources, adjust the load state, and realize the economic operation of the power grid. Based on the MATLAB platform, a group model of MG is built and simulated. The results show the effectiveness of the proposed control method. Compared with other methods, the proposed control method has higher income and minimum carbon emission and realizes the economic and environmental protection system operation.

## 1. Introduction

With the continuous advancement of new energy power generation technology, communication technology, Internet technology, and other power industry technologies and new-generation information and communication technologies, the IoT technology and the distribution network are deeply integrated to form the Internet of distribution things and microgrids (MG). The use of networked supply and intelligent management technology can play a series role between the user side and distributed energy [1, 2]. With the continuous development of power grid technology, MG is a relatively independent system, which can not only operate independently but also constitute a multi-energy complementary intelligent MG group. Among them, the shortcomings of the intermittent power output of distributed power can be compensated by reasonable regulation, so as to ensure the quality and reliability of the power supply [3].

At present, MG still faces greater challenges in regulating distributed generation (DG), battery energy storage system (BESS) equipment, and loads [4, 5]. Ref. [6] proposes a “source-storage-load” coordination balance algorithm based on deep learning, which enables the system and user load to achieve Nash equilibrium without prior information, and optimizes the MG’s intelligent control capabilities. With the development of research on the mobile BESS characteristics of EV, it enters the MG as a special DG [7]. Ref. [8] constructed a real-time MG optimal energy management system, by using the random forest method to predict the EV driving mode to schedule the charging and discharging of the EV battery, which not only improves the consumption of distributed energy but also improves its utilization efficiency. Ref. [9] studies the energy management framework of intelligent MG and analyzes the energy optimization among household load, EV, BESS, and distribution network. Most of the control strategies proposed in the above literature are from the perspective of MG and use demand response to guide users



to optimize the MG economy [10]. However, the uncertainty of EV and collaborative optimization of distributed energy in the MG group of power distribution IoT still need to be further studied.

Therefore, under the framework of cloud edge collaborative in power distribution IoT, a MG regulation method based on deep learning is proposed. Based on the established MG system model, as well as the system optimization objectives and constraints, the edge side training learning of the deep learning algorithm is used to regulate and control the MG group.

## 2. MG System Model in Power Distribution IoT

Combining edge computing with cloud computing, the cloud edge collaborative computing framework is constructed, and the power distribution IoT architecture based on cloud edge collaboration is established, as shown in Figure 1. Taking the edge computing group as the basic unit, according to the logic structure of cooperative autonomy between groups and cloud edge collaborative control, the mathematical model and training learning model of MG group computing are established; finally, the control optimization calculation of the MG group is carried out on the power distribution cloud platform.

Among them, the end device mainly collects the data of each MG for modeling; the edge node has the edge computing ability, collects the data of the end device and determines the optimization objectives and constraints, and trains and learns the data model based on the deep learning algorithm; the power distribution cloud platform uses the data information of each edge node and considers the target optimization model to achieve a larger scale. The optimal energy distribution of the MG group is proposed.

The topology structure of MG is shown in Figure 2, which mainly consists of wind turbine (WT), photovoltaic (PV) energy, BESS, gas generator, EV, fuel cell, energy conversion device, and users.

In the MG system, MG is connected with the main network, and vehicle to grid is introduced. The role of vehicle to grid is to stimulate the charging of the vacant EV, so that it does not need to be charged during the peak load, which reduces the power supply pressure of the main network, and the electric energy stored in the EV can be sent to the main network, increasing the power supply in the system. As the power supply of MG, DG will change under the influence of weather and other factors, and the system will be adjusted accordingly. The power in the system eventually flows to the user.

*2.1. Generating Unit Side.* Renewable energy such as WT power generation and PV power generation is increasingly widely used in MG. At the same time, BESS can effectively solve the problem of intermittent output of distributed energy [11]. Therefore, the MG group adopts the WT/optical/storage/grid collaborative power generation mode.

The output of WT is closely related to environmental wind speed, wind cut-in and cut-out speed, and rated wind speed. PV output power  $P_{PV}$  is determined by the output

power of PV modules, solar irradiance, and ambient temperature under standard conditions. The battery next state of charge (SOC( $i + 1$ )) is related to the current state of battery (SOC( $i$ )).

*2.2. End Side EV Model.* The randomness of EV is mainly reflected in the uncertainty of the time to access/leave the MG and the randomness of the initial SOC due to the driving distance. The end time and mileage of EV generally follow normal distribution [12]. Therefore, based on the Monte Carlo algorithm, the probability model of end time and mileage of EV is established, which is expressed as follows:

$$f(t) = \begin{cases} \frac{1}{t\sigma_t\sqrt{2\pi}} \exp\left[-\frac{(t+24-\mu_t)^2}{2\sigma_t^2}\right], & 0 \leq t \leq \mu_t - 12, \\ \frac{1}{t\sigma_t\sqrt{2\pi}} \exp\left[-\frac{(t-\mu_t)^2}{2\sigma_t^2}\right], & \mu_t - 12 < t \leq 24, \end{cases}$$

$$f(d) = \frac{1}{t\sigma_d\sqrt{2\pi}} \exp\left[-\frac{(\text{Ind} - \mu_d)^2}{2\sigma_d^2}\right], \quad 0 < d < 120, \quad (1)$$

where  $t$  is the end time of driving,  $\sigma_t = 3.41$ ,  $\mu_t = 17.47$ ,  $d$  is the mileage,  $\sigma_d = 3.24$ , and  $\mu_d = 8.92$ .

EV in MG can be divided into dispatchable vehicles and schedulable vehicles according to whether the owners agree to participate in centralized control. Among them, disorderly charging is adopted for nonschedulable vehicles; that is, the owners charge by returning time and driving demand of EV in the next period; and orderly charging is adopted for schedulable vehicles; that is, under the time of use price mechanism, the owners can charge uniformly within the specified time [13, 14].

*2.3. End User Load Unit.* Based on the comparative analysis of users' usage habits and load types, user loads can be divided into base load, reducible load, translatable load, and interruptible load [15, 16]. Among them, the base load is a necessary load and does not have the ability to adjust. The latter three are adjustable loads, which can be adjusted according to electricity price or other incentive policies. Electric water heater (EWH) and air conditioning (AC) are widely used and have BESS characteristics. Optimization strategies can be adopted to control their output power in peak power consumption as a representative of translatable load; its working range can be adjusted to the low power consumption period [17]. The user side of the MG system can adjust the load utilization through the electricity price mechanism.

## 3. Optimization Model and Control Strategy of MG Group in Power Distribution IoT

Under the time-of-use price mechanism, the overall load demand of users in the MG group of the power distribution IoT will inevitably change [18, 19]. Therefore, under the cloud-side collaborative architecture, deep learning is used to control the MG group, rationally regulate the coordinated

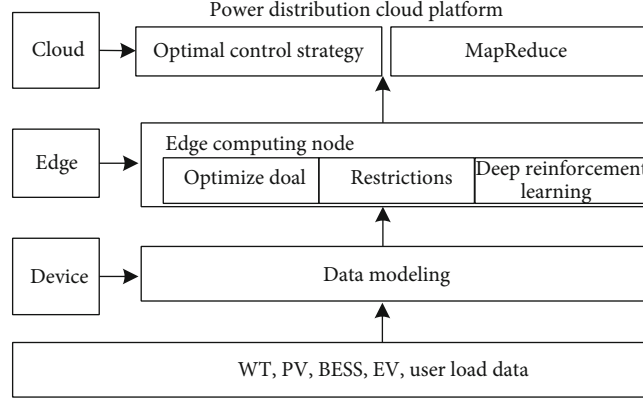


FIGURE 1: Power distribution IoT system architecture based on “cloud-edge-end.”

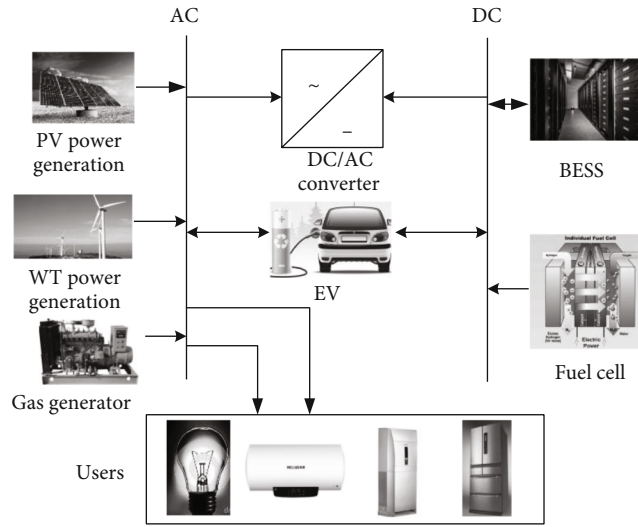


FIGURE 2: Topology of MG system.

output of multiple energy sources, adjust the load status, and realize the economic operation of the power grid.

### 3.1. The Optimization Goal of Cloud-Side Collaboration

**3.1.1. Daily Operating Cost.** The operating cost of a MG group in a cycle is an important factor to improve the economic benefits of users, including its initial investment cost, daily operation and maintenance costs, and load transfer compensation after users participate in the time-of-use electricity price mechanism [20, 21]. The optimization objective function is

$$\min C_{\text{lower}} = \Delta i \sum_{i=1}^{24} [C_{\text{WT}} + C_{\text{PV}} + C_{\text{BESS}} + C_G], \quad (2)$$

where  $C_{\text{WT}}$ ,  $C_{\text{PV}}$ , and  $C_{\text{BESS}}$  are the total operating costs of WT, PV array, and BESS, respectively, and  $C_G$  is the interactive power cost of MG and large grid.

**3.1.2. Heterogeneous Energy Synergy and Charging Power Optimization.** In the low-voltage distribution network, the sources of charging energy usually include BESS, WT, and

PV. However, since PV and WT are greatly affected by environmental factors, if the BESS can be used to balance the impact of environmental factors, the fluctuation of charging power in the MG can be reduced [22, 23]. The calculation formula of charging power is as follows:

$$\begin{aligned} \min & \left( P_{\text{charge}}^t p_e^t + P_{\text{WT}}^t + P_{\text{PV}}^t \right) \\ P_{\text{charge}}^t &= P_{\text{BESS}}^t + P_{\text{WT}}^t + P_{\text{PV}}^t, \end{aligned} \quad (3)$$

where  $t$  is the time,  $P_{\text{BESS}}^t$  is the charging power of BESS,  $P_{\text{WT}}^t$  is the charging power of WT,  $P_{\text{PV}}^t$  is the charging power of PV,  $P_{\text{charge}}^t$  is the total charging power of MG with upper and lower limits.  $p_e^t$  is the cost price, and  $P_{\text{WT}}^t$  and  $P_{\text{PV}}^t$  are the cost of WT power generation and PV power generation, respectively.

### 3.2. Constraint Condition

**3.2.1. EV Constraints.** The constraints of EV should not only consider the characteristics of EV but also meet the normal needs of car owners. The state of charge constraints of EV

and the charging and discharging power constraints of EV are as follows:

$$\begin{aligned} \text{SOC}_{\text{EV},\min} &\leq \text{SOC}_{\text{EV}}(i) \leq \text{SOC}_{\text{EV},\max}, \\ P_{\text{EV},\text{dis}} &\leq P_c(i) \leq P_{\text{EV},\text{cha}}, \end{aligned} \quad (4)$$

where  $\text{SOC}_{\text{EV},\max}$  is 0.95,  $\text{SOC}_{\text{EV},\min}$  is 0.2,  $P_{\text{EV},\text{dis}}$  is the maximum discharge power of EV, and  $P_{\text{EV},\text{cha}}$  is the maximum charging power.

**3.2.2. Supply and Demand Balance Constraints of MG.** In order to ensure the normal life of users, the power provided by MG should be balanced with the power required by users:

$$P_{\text{load}}(i) - P_{\text{WT}}(i) - P_{\text{PV}}(i) - P_{\text{BESS}}(i) - P_G(i) = 0, \quad (5)$$

where  $P_{\text{load}}(i)$  is the load demand at  $i$  time after MG participates in the control strategy.

**3.3. Edge-Side Training Learning Based on Deep Learning Algorithm.** By combining the edge computing capabilities of edge nodes with the super perception of deep learning and the decision-making of reinforcement learning, the deep reinforcement algorithm can perform output control based on the analysis of input data, making it closer to the way people think [24, 25]. Reinforcement learning is based on the Markov decision process (MDP), which makes the transition of the system at the next moment independent of the previous moment [26, 27]. The deep learning algorithm uses the following function value update method to approximate the Q function:

$$\varphi_{t+1} = \varphi_t + \alpha \left( r_{t+1} + \gamma \max_a Q(s_{t+1}, a, \varphi) Q(s_t, a_t, \varphi) \right) - \nabla_{\varphi} Q(s_t, a_t, \varphi), \quad (6)$$

where  $\alpha$  is the learning rate,  $\varphi$  is the neural network weight,  $\gamma$  is the discount factor,  $s$  is the system state, and  $a$  is the action strategy, by which  $\alpha = 1$ ,  $\gamma = 0.85$ . When training a neural network, use the mean square error to define the error function:

$$L(\varphi) = E \left[ \left( r_{t+1} + \gamma \max_a Q(s_{t+1}, a_{t+1}, \varphi) - Q(s_t, a_t, \varphi) \right)^2 \right]. \quad (7)$$

Obtain the gradient of the error function in the  $\varphi$  direction, update the parameters by means of stochastic gradient descent, and obtain the optimal strategy on the basis of obtaining the optimal Q value. In the deep learning training process, if the selection action and the evaluation action come from the same Q value of the same network, the final result may have a large error due to overestimation. The dual deep learning calculates the maximum Q value in the main network for selection actions, and the target Q value calculation is performed in the target network, as shown in the fol-

lowing formula:

$$Q_t^{\text{DoubleQ}} = r_{t+1} + \gamma Q \left( s_{t+1}, \arg \max_a Q(s_{t+1}, a; \varphi_t); \varphi_t^- \right). \quad (8)$$

In order to alleviate the problem of model overestimation, the model usually needs to control a small difference range between the target Q value and the actual Q value difference, and this helps to improve the algorithm convergence speed.

**3.4. Control Strategy of MG Group Based on Power Distribution Cloud Platform.** The MG group control strategy takes BESS, EV, gas storage, and time as system states, discretizes the original continuous MG operation process, and separates charging and discharging and other forms of electrical energy [28, 29]. As the action strategy, assume that the current state is  $s_t$ , the next state is  $s_{t+1}$ , the allowed action strategy is  $a$ , and the action process includes changes in equivalent parameters.

**3.4.1. MDP Tuple Description.** The state space  $s$  consists of three parts: controllable battery  $s_b$ , uncontrollable PV and load  $s_{\text{PV},l}$ , and time series  $s_m$ :

$$s = s_b \times s_{\text{PV},l} \times s_m. \quad (9)$$

The reward function is a real-time reward function, which is aimed at evaluating a point in time information, and cannot explain the quality of the overall strategy. Therefore, it is necessary to define the state action value function to represent the long-term effect of the strategy on the state:

$$Q_h(s, a) = E_h \left[ \sum_{t=0}^{T-1} \gamma^t r | s_t = s, a_t = a \right]. \quad (10)$$

**3.4.2. Control Strategy of MG Cluster.** In the power distribution cloud platform system, the state input includes BESS battery storage capacity  $E$ , natural gas storage capacity  $G$ , EV storage capacity  $V$ , and time  $t$ . The discrete-time state quantity is 48. Different state variables have different ways to determine the action strategy  $a$  in MG. After the input state and action strategy are determined, the online learning can be synchronized [30]. The multiple iterations of the Q algorithm can make the Q value table tend to converge, so as to determine the optimal scheduling route. The overall flow of the algorithm is shown in Figure 3.

The calculation of each state conversion income usually includes variable information such as selection environment, current time price, and natural gas price and then fills in the  $R$  matrix of the corresponding action under the state [30]. If there is no action corresponding to the state in  $R$ , the  $R$  value table is generated. In state  $s_t$ , according to the BESS, natural gas storage, automobile power storage, and current time contained in the current MG group, determine the action  $A_t$  that should be taken; then, the system will enter the next state.

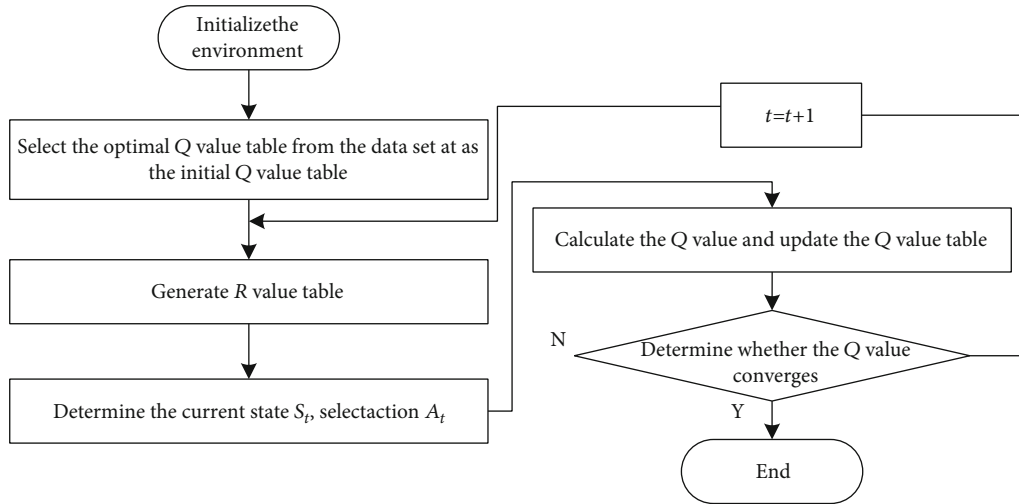


FIGURE 3: Control strategy of double level optimization model.

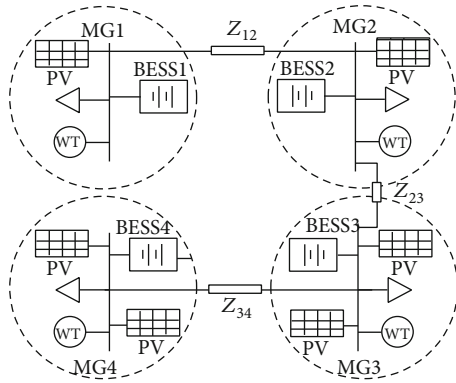


FIGURE 4: Structure of MG group.

#### 4. Simulation Results and Analysis

In order to verify the effectiveness of the proposed MG group control method, a group model containing 4 autonomous MG was built in the real-time simulation platform MATLAB, and its topology is shown in Figure 4.

The rated voltage/frequency of the MG group is 380 V/50 Hz. The MG group includes PV module unit, WT power generation unit, and BESS, and the specific capacity is 400 kW, 600 kW, and 400 kW, respectively. According to the actual load demand of a city in China, the time-of-the-art price mechanism is adopted for electricity sales and purchase in MG. According to the actual load demand of a city in China, the valley load period is from 22:00 to 5:00 the next day; 6:00 to 7:00, 11:00 to 12:00, and 17:00 to 18:00 are load sharing periods; the rest are peak load periods. The electricity purchase prices of peak, flat valley, and valley are 0.83 yuan/kWh, 0.49 yuan/kWh, and 0.17 yuan/kWh, respectively, and the electricity selling prices are 0.65 yuan/kWh, 0.38 yuan/kWh, and 0.13 yuan/kWh, respectively.

**4.1. Regulation Results of Single MG.** The time-of-use electricity price mechanism is used to guide users to adjust the usage habits of adjustable loads in order to achieve the purpose of

“peak cutting and valley filling.” The overall load curve before and after optimization of the MG is shown in Figure 5.

It can be seen from Figure 5 that before optimization, the overall load curve of consumers fluctuates greatly, and the peak value of electricity consumption is concentrated in the period of high electricity price. After the user load participates in the control strategy, the overall load curve changes, showing that the daytime demand power decreases, while the night time demand power increases, and the load decreases during the peak period, thus reducing the peak valley difference and smoothing the load curve. It can be seen that after the energy regulation of MG, the total load energy consumption is reduced and the economy of the system is improved.

**4.2. Optimization Results of MG Group.** By adjusting the optimal coordination mode of distributed energy, BESS, and load, the capacity utilization rate of the MG group can be improved, and the economic benefit can be improved. The results of heterogeneous energy optimization control are shown in Figure 6.

It can be seen from Figure 6 that during the period from 22:00 to 06:00 the next day, the WT of distributed energy has a large output. Under the condition of ensuring the normal load demand, the BESS charges. Since the power generated by distributed energy is greater than the load demand, the MG sells electricity to other loads in the MG group, so as to increase the economic benefits of users. However, during 11:00-16:00, during the peak period of MG power consumption, the output of WT is reduced, and the output of PV power generation is large. At the same time, due to the peak electricity price period, the BESS starts to cooperate with the WT and PV array to output at the same time, so as to reduce the consumers’ purchase of electricity from the large grid; at the same time, the BESS stores the energy during the low electricity price period and when the distributed energy output has surplus. In the high electricity price period, it not only ensures the stability of MG power supply but also improves the consumption capacity of distributed energy,

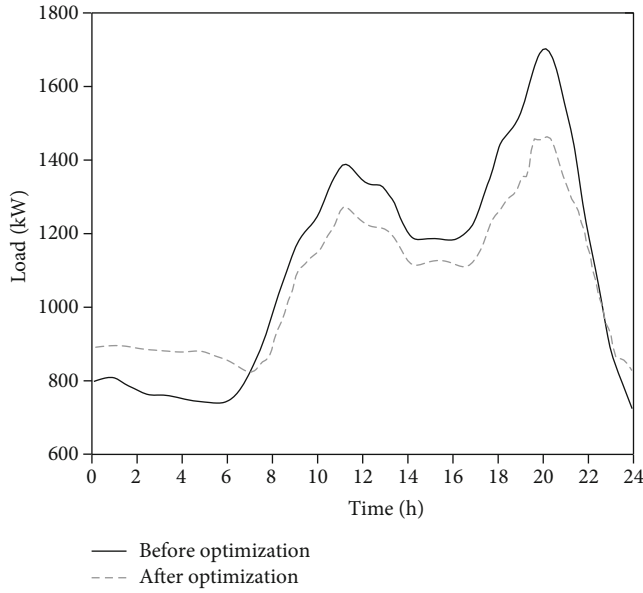


FIGURE 5: Overall load curve before and after MG optimization.

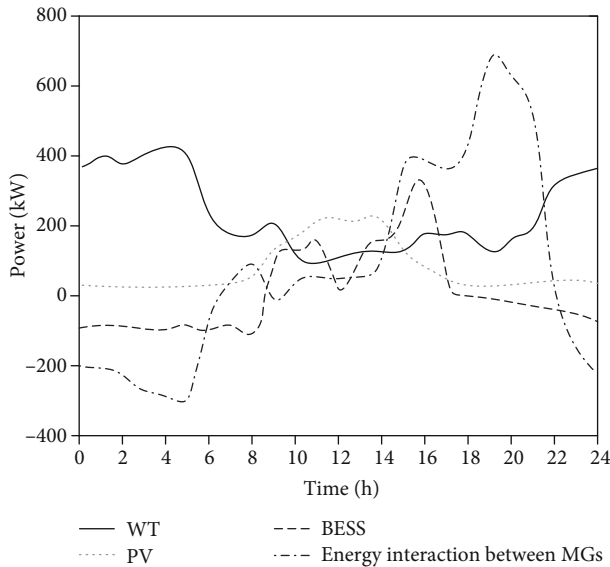


FIGURE 6: Daily power curve of multienergy complementary in MG group.

thus smoothing the overall load curve and achieving the purpose of peak load shifting and valley filling.

**4.3. Comparative Analysis of Different Methods.** In order to demonstrate the economic and environmental protection of the proposed method, it is compared with the methods in Refs. [6, 8, 9]. Among them, the economy and environmental protection are calculated quantitatively from the electricity purchase cost and carbon emission of MG, respectively. The product of the two is used as the evaluation index. The smaller the value is, the stronger the regulation ability is. The experimental results of the growth trend of economic

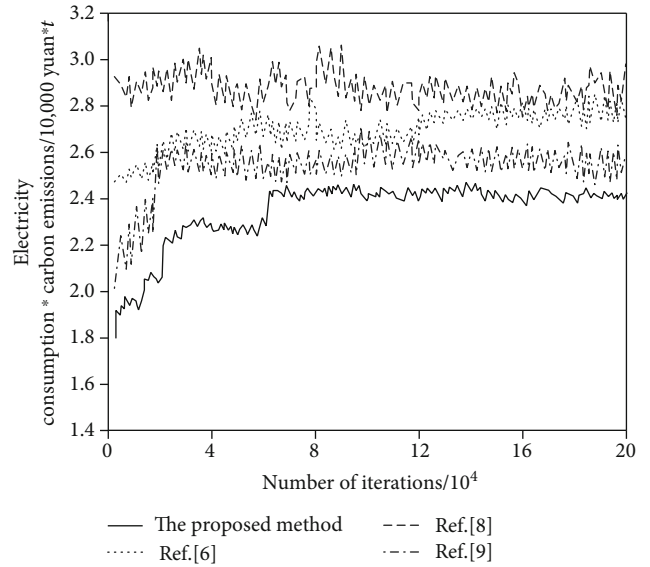


FIGURE 7: Growth trend of total economic income in MG group.

and environmental protection of the MG group are shown in Figure 7.

As can be seen from Figure 7, at the beginning of the iteration, the performance of each algorithm is low, but with the increase of the number of iterations, the optimal solution is constantly approaching and finally tends to converge; the economic and environmental performance is optimal, which is about 24000 yuan \* t, exceeding the benefit of the empirical learning algorithm. The algorithm used in Refs. [6, 8] has a relatively small amount of calculation, so it can converge quickly, but the cost of purchasing electricity is very high, about 26500 yuan, and the overall regulation performance is poor. Ref. [9] grows rapidly in the initial iteration stage, but due to the lack of prediction for future strategies, it grows slowly in the later stage and has poor performance. It can be demonstrated that the proposed method has a good ability of energy coordination and optimization. Through the improved deep reinforcement learning algorithm control strategy, the economic and environmental protection of the system has been greatly improved.

## 5. Conclusions

With the promotion of distributed energy, the number of MG has increased dramatically, forming MG groups. In order to improve the coordination and optimization of MG group energy, a control strategy based on deep reinforcement learning is proposed. Based on the cloud-side collaborative power distribution IoT architecture, the system model of the MG is proposed and interconnected to construct the system architecture of the MG group. In addition, the edge-side training and learning of the deep reinforcement learning algorithm are used to control the MG group, rationally regulate the coordinated optimization of multiple energy sources, and realize the economic and environmental protection operation of the MG group. A MG group model was built on the MATLAB platform to conduct simulation

experiments. The results show that the proposed method introduces the time-of-use electricity price mechanism to regulate load operation and achieve the purpose of peak shaving and valley filling, and the overall energy consumption is small, and the economic performance is better. Compared with other methods, the system has the smallest carbon emissions, maximizes the consumption of renewable energy, and realizes economical and environmentally friendly system operation.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare no conflicts of interest.

## Acknowledgments

This study was financially supported by the National Key Research and Development Project (No.2018YFB0905000), QingHai Department of Science and Technology (No.2018-GX-A6), and Shandong Province Graduate Education Quality Improvement Program (No.SDYKC19103).

## References

- [1] L. Chang, Z. Jiankun, Z. Dongming et al., "Review on the research of flexible and safe operation of renewable energy microgrids using energy storage systems," *Proceedings of the CSEE*, vol. 37, no. 4, pp. 1–18, 2019.
- [2] S. Yi, P. Junyi, and J. Dongsheng, "Smart home appliance control strategy considering user behavior uncertainty," *Power System Protection and Control*, vol. 46, no. 17, pp. 109–117, 2018.
- [3] J. Zhu, X. Mo, T. Zhu, T. Luo, M. Liu, and Y. Guo, "Real-time stochastic operation strategy of a microgrid using approximate dynamic programming-based spatiotemporal decomposition approach," *IET Renewable Power Generation*, vol. 13, no. 16, pp. 3061–3070, 2019.
- [4] L. Daxing, X. Gefei, L. Wenlong, Y. Baoxin, and X. Lai, "Energy management system for smart home," *Proceedings of the CSU-EPSA*, vol. 28, Supplement 1, pp. 186–193, 2016.
- [5] L. Han, A. T. Eseye, Z. Jianhua, and Z. dehua, "Optimal energy management for industrial microgrids with high penetration renewables," *Protection and Control of Modern Power Systems*, vol. 2, no. 2, pp. 122–135, 2017.
- [6] P. Zeng, H. Li, H. He, and S. Li, "Dynamic energy management of a microgrid using approximate dynamic programming and deep recurrent neural network learning," *IEEE Transactions on Smart Grid*, vol. 10, no. 4, pp. 4435–4445, 2019.
- [7] Z. Liu, Y. Luo, R. Zhuo, and X. Jin, "Distributed reinforcement learning to coordinate current sharing and voltage restoration for islanded DC microgrid," *Journal of Modern Power Systems and Clean Energy*, vol. 6, no. 2, pp. 364–374, 2018.
- [8] E. Foruzan, L.-K. Soh, and S. Asgarpoor, "Reinforcement learning approach for optimal distributed energy management in a microgrid," *IEEE Transactions on Power Systems*, vol. 33, no. 5, pp. 5749–5758, 2018.
- [9] Z. Yanyu, Z. Peng, and Z. Chuazhi, "Optimal scheduling algorithm for home energy management system in smart grid environment," *Power System Protection and Control*, vol. 44, no. 4, pp. 18–26, 2016.
- [10] L. Lei and H. Sipeng, "Hierarchical and optimized operation strategy of active distribution network with microgrid," *Electrical Measurement & Instrumentation*, vol. 56, no. 20, pp. 76–81, 2019.
- [11] Z. Qiaolai, M. Q. Yundong, and W. Yu, "On microgrid interconnection research based on demand side response," *Electrical Measurement & Instrumentation*, vol. 56, no. 10, pp. 34–39, 2019.
- [12] V. S. Tabar, M. A. Jirdehi, and R. Hemmati, "Energy management in microgrid based on the multi objective stochastic programming incorporating portable renewable energy resource as demand response option," *Energy*, vol. 118, pp. 827–839, 2017.
- [13] P. D. Lund, J. Lindgren, J. Mikkola, and J. Salpakari, "Review of energy system flexibility measures to enable high levels of variable renewable electricity," *Renewable and Sustainable Energy Reviews*, vol. 45, no. 45, pp. 785–807, 2015.
- [14] M. Muratori, "Impact of uncoordinated plug-in electric vehicle charging on residential power demand," *Nature Energy*, vol. 3, no. 3, pp. 193–201, 2018.
- [15] M. Zeng, G. Wu, H. J. Wang, R. Li, B. Zeng, and C. J. Sun, "Resident demand-side response control strategy considering user satisfaction in the context of smart grid," *Power System Technology*, vol. 40, no. 10, pp. 2917–2923, 2016.
- [16] H. Feng, Z. Chengbi, and M. Hong, "Load priority control of household microgrid based on time-of-use electricity price and demand response," *Electrical Measurement & Instrumentation*, vol. 55, no. 19, pp. 41–45, 2018.
- [17] C. Jindong, W. Shengwen, and X. Yechun, "Research on the smart grid data management technology framework from the perspective of block alliance chain," *Chinese Journal of Electrical Engineering*, vol. 40, no. 3, pp. 836–848, 2020.
- [18] X. Y. Zhao, S. Wang, X. H. Wu, and J. Liu, "Coordinated control strategy research of micro-grid including distributed generations and electric vehicles," *Power System Technology*, vol. 40, no. 12, pp. 3732–3740, 2016.
- [19] N. Wu and H. Wang, "Deep learning adaptive dynamic programming for real time energy management and control strategy of micro-grid," *Journal of Cleaner Production*, vol. 204, no. 1, pp. 1169–1177, 2018.
- [20] M. Afrasiabi, M. Mohammadi, M. Rastegar, and A. Kargarian, "Multi-agent microgrid energy management based on deep learning forecaster," *Energy*, vol. 186, no. 11, article 115873, 2019.
- [21] L. Wen, K. Zhou, S. Yang, and X. Lu, "Optimal load dispatch of community microgrid with deep learning based solar power and load forecasting," *Energy*, vol. 171, no. 3, pp. 1053–1065, 2019.
- [22] G. Y. Lee, B. S. Ko, J. Cho, and R. Y. Kim, "A distributed control method based on a voltage sensitivity matrix in DC microgrids with low-speed communication," *IEEE Transactions on Smart Grid*, vol. 10, no. 4, pp. 3809–3817, 2019.
- [23] Y. Du and F. Li, "Intelligent multi-microgrid energy management based on deep neural network and model-free reinforcement learning," *IEEE Transactions on Smart Grid*, vol. 11, no. 2, pp. 1066–1076, 2020.
- [24] D. S. Rwegasira, I. S. Ben Dhaou, A. Kondoro et al., "A demand-response scheme using multi-agent system for smart

- DC microgrid,” *International Journal of Embedded and Real-Time Communication Systems*, vol. 10, no. 1, pp. 48–68, 2019.
- [25] X. Lu, X. Xiao, L. Xiao, C. Dai, M. Peng, and H. V. Poor, “Reinforcement learning-based microgrid energy trading with a reduced power plant schedule,” *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 10728–10737, 2019.
- [26] Z. Wang, W. Wu, and B. Zhang, “A distributed quasi-Newton method for droop-free primary frequency control in autonomous microgrids,” *IEEE Transactions on Smart Grid*, vol. 9, no. 3, pp. 2214–2223, 2018.
- [27] Q. Wang, Y. Liu, W. Song, and K. Xuan, “Improved dynamic control method for energy storage units in PV dominated microgrids,” *Archives of Electrical Engineering*, vol. 67, no. 4, pp. 885–898, 2018.
- [28] J. He, Y. Pan, B. Liang, and C. Wang, “A simple decentralized islanding microgrid power sharing method without using droop control,” *IEEE Transactions on Smart Grid*, vol. 9, no. 6, pp. 6128–6139, 2018.
- [29] J. O. Lee, Y. S. Kim, and S. I. Moon, “Novel supervisory control method for islanded droop-based AC/DC microgrids,” *IEEE Transactions on Power Systems*, vol. 34, no. 3, pp. 2140–2151, 2019.
- [30] Q. Li, M. Gao, H. Lin, Z. Chen, and M. Chen, “MAS-based distributed control method for multi-microgrids with high-penetration renewable energy,” *Energy*, vol. 171, no. 3, pp. 284–295, 2019.

## Research Article

# An Empirical Study on Optimal the Allocations in Advertising and Operation Innovation on Supply Chain Alliance for Complex Data Analysis

Jiang-Tao Wang <sup>1</sup>, Jian-Jun Yu,<sup>2</sup> Yu-Hsi Yuan <sup>3</sup>, Sang-Bing Tsai <sup>4</sup>,  
and Shu-Fen Zhang <sup>5</sup>

<sup>1</sup>Zhongshan Institute, University of Electronic Science and Technology of China, Zhongshan 528400, China

<sup>2</sup>School of Business Administration, South China University of Technology, Guangzhou 510641, China

<sup>3</sup>Department of Labor & Human Resources, Chinese Culture University, Taipei 111, Taiwan

<sup>4</sup>Regional Green Economy Development Research Center, School of Business, WUYI University, Wuyishan 354300, China

<sup>5</sup>School of Chinese Medicine, Guangdong Pharmaceutical University, Guangzhou 510006, China

Correspondence should be addressed to Yu-Hsi Yuan; [yuanyh@gm.ypu.edu.tw](mailto:yuanyh@gm.ypu.edu.tw) and Shu-Fen Zhang; [zsf063@gdpu.edu.cn](mailto:zsf063@gdpu.edu.cn)

Received 15 December 2020; Revised 11 January 2021; Accepted 28 January 2021; Published 25 February 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Jiang-Tao Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Effective and efficient closed-loop supply chain processes can provide a significant competitive edge for companies. This study considered three investment strategies in the process of initiating closed-loop supply chain alliances. The results showed that a promised proportion has a significant effect on investment decisions under a pure investment strategy. Furthermore, a reasonable promised proportion can coordinate the supply chain under a pure innovation strategy but cannot in a pure advertising strategy. Upstream (i.e., innovation) investments decrease wholesale and retail prices, while downstream ones increase retail and wholesale prices. Increasing innovation investment can transform benefits to the downstream, while increasing advertising investment may cause opportunism. A hybrid investment strategy balances upstream and downstream investment simultaneously and provides insights into optimizing the supply chain system in investments.

## 1. Introduction

Current societies and enterprises are paying increasing attention to environmental pollution. Thus, more attention is being paid to product reuse management. Some types of government legislation require manufacturers to deal with their end-of-life products and waste production. Many manufacturing firms have thus begun to focus on product recovery management, including returns, refurbishing, recycling, remanufacturing, and marketing, to comply with rigid environmental regulations [1]. Product recycling can in fact benefit a company if it is handled properly. Dell reported that since 2014, it has recycled 4.2 million pounds, reduced its product carbon footprint by 11%, and achieved cost savings through its recycling supply chain. Meanwhile, 14 auto part remanufacturing enterprises and 35 home appliance manufacturers in China, which

had been identified as national e-waste collection and recycling pilot projects, have saved almost 155 billion RMB in environmental benefits per year [2, 3]. Hence, closed-loop supply chain issues have gained considerable attention among both academia and practitioners due to the positive environmental effects [4–6] and economic benefits [7].

In practice, it is an effective way to implement closed-loop supply chains with suitable partners to copy with rigid environment regulations and enjoy the economic benefits. Hence, this study focused on building the closed-loop supply chains via upstream and downstream investment strategies.

The extant research on closed-loop supply chain issues has mainly focused on designing, planning [8–10], and surveying [11]. For a review of reverse logistics and closed-loop supply chain literature, refer to Guide & Wassenhove [12], Agrawal et al. [13], Govindan et al. [4] and Islam & Huda [8].



Reverse-logistic management issues in traditional supply chain management refer to remanufactured products [14, 15]. A complete literature review in this field can be found in Fleischmann et al. [16], Guide et al. [17], Souza [18], Stindt & Sahamie [19] and Marić & Opazo-Basáez (2019). This stream of literature has focused on network design [20, 21], inventory control [22], reverse channel structure [23, 24], simulation (Abid et al., 2019), price, and coordination [25, 26]. In practice, operational innovation and advertising are both often used to build a company's competitive advantage, i.e., advertisement aims to stimulate demand or opens up the sales market, while operational innovation can improve a company's performance through reducing costs. However, these two investment decisions occur at the front and back ends of the supply chain, respectively. Advertising decisions and operational innovation decisions are frequently encountered and widely discussed topics in operations management. Inspired by such research, this study adopted a similar approach to consider the problems of upstream and downstream investment allocation in the supply chain.

Advertising can build a stock of goodwill and promotes product sales. Cooperative advertising is a coordinating mechanism in the marketing channel that is widely used in supply chain management. For example, coop advertising was first adopted to examine promotion effect problems in supply chains (Berge, 1973). Most of the literature has focused on evaluating the impact of advertising investment on market demand (Dai & Chao, 2013; [27, 28]). Consistent with these assumptions, we assumed that advertising investment could improve the volume of sales, and then we considered advertising investment decisions in supply chains.

Operational innovation in supply chains pertains to improvements in quality, yield, delivery time, and supply cost [29]. Little attention was paid to the recycling and reuse of waste products in the early stage, leading to low reproduction efficiency. With the increasing emphasis on product recycling, more and more research has suggested that it is necessary to improve the processes for recycled products via innovation investment [30]. A comprehensive performance measurement system was developed to measure the performance of firms with respect to innovation policy and marketing strategy [31]. Most researchers have realized that innovation investment can result in either cost reduction or quality improvement [31–33]. Based on the relations between innovation and product cost found in the literature [33], we considered innovation investment decisions in the supply chain.

Note that the two streams of literature mentioned above either considered advertising investment or innovation investment in closed-loop supply chains. However, a closed-loop supply chain alliance can simultaneously consider innovation investment in the upstream and advertising investment in the downstream to enjoy the economic benefits. How to allocate investments on advertising and operational innovation is one of challenges of closed-loop supply chain systems. Any initiator of a closed-loop supply chain needs to weigh investment decisions on upstream and downstream simultaneously in the supply chain. Hence, this study focus on the allocations on/between advertising and operation innovation on supply chain, and investigates the optimal investment allocation on/between the two decisions within a supply chain system.

The rest of this paper is organized as follows. Section 1 presents the model descriptions. The model analysis will examine the investment decision under pure strategies as well as under mixed strategies in Section 2. Section 3 presents numerical results based on the theoretical results. A summary of this research and discuss future research directions are conducted in Section 4. Some proofs are presented in the supplementary materials (available here).

## 2. Model Description

Consider a supply chain system composed of a manufacturer and a seller, in which the manufacturer, possessing manufacturing/remanufacturing production lines, can directly produce products with new components and remanufacture products with used products; it can then distribute new products and recover used products through the seller. Consistent with Xu et al. [33], one returned product can be remanufactured into one remanufactured product. The original remanufacture cost is  $c$  and can be reduced to  $c - \beta x$  ( $0 < \beta \ll 1$ ) with investing operational innovation  $x^2$  to improve the remanufacturing production lines [32]. Under a price-sensitive market, sales in the market are  $Q = D - kp$  and can increase to  $Q + \alpha y$  ( $0 < \beta \ll \alpha < 1, \alpha < k$ ) as advertising investment increases to  $y^2$  [32, 34, 35], where the potential initial market is  $D$ , and the retail price is  $p$ .

There are two scenarios for building the supply chain alliance. A leading manufacturing enterprise may invite a seller to join the alliance and promise to bear part of the advertising investment for the retailer to alleviate the seller's investment risk. A seller may actively request to join the alliance and promise to bear part of the innovation investment for improving technology to reduce reproduction costs and then hope to decrease the wholesale price for itself. Then, the supply chain alliance must make allocation decisions in investments in advertising and innovation to optimize alliance performance.

## 3. Model Analysis

*3.1. Pure Innovation Investment Strategy.* A supply chain center uses an innovation investment strategy to reduce remanufacturing costs. Under a price-sensitive market, the central planner makes the innovation investment  $x^2$  as

$$\pi^{SI}(p, x) = (p - c + \beta x)(D - kp) - x^2. \quad (1)$$

**Lemma 1.** *A pure innovation investment strategy has optimal innovation investment and retail price as follows:*

$$p^{SI*} = \frac{2D + 2kc - Dk\beta^2}{k(4 - k\beta^2)}, x^{SI*} = \frac{\beta(D - kc)}{4 - k\beta^2}. \quad (2)$$

To build the supply chain alliance, the seller promises to undertake part of the investment to improve the remanufacturing production line. Then, the decision-making sequence is as follows: (1) The seller promises to undertake a proportion of the investment (i.e.,  $t$ ) and becomes responsible for

recycling used products in advance. (2) The manufacturer then determines innovation investment  $x^2$  and wholesale price  $w$ . (3) The seller finally determines the retail price in the price-sensitive market.

The seller's decision equation is as follows:

$$\pi_r^{SI}(p | x, w, t) = (p - w)(D - kp) - tx^2. \quad (3)$$

The manufacturer's decision equation is as follows:

$$\pi_m^{SI}(w, x | p, Q) = (w - c + \beta x)(D - kp) - (1 - t)x^2. \quad (4)$$

**Lemma 2.** *Under a pure innovation investment strategy, the supply chain alliance exists in equilibrium for both the manufacturer and seller as follows:*

$$\begin{aligned} p_t^{SI*} &= \frac{2ck(-1+t) + D(-6 + \beta^2k + 6t)}{k(-8 + \beta^2k + 8t)}, \\ Q_t^{SI*} &= \frac{2(D - ck)(-1+t)}{-8 + \beta^2k + 8t}, \\ x_t^{SI*} &= \frac{\beta(-D + ck)}{-8 + \beta^2k + 8t}, \\ w_t^{SI*} &= \frac{4ck(-1+t) + D(-4 + \beta^2k + 4t)}{k(-8 + \beta^2k + 8t)}. \end{aligned} \quad (5)$$

The equilibrium suggests that manufacturer would made operational innovations  $x_t^{SI*}$  and then charged the wholesale price  $w_t^{SI*}$  with the promised undertaking proportion  $t$ .

**Proposition 3.** *Under a pure innovation investment strategy, the promised proportion has a significant effect on the innovation investment. Specifically,  $x^{SI*} \geq x_t^{SI*}$  with  $t \leq 1/2$ , and  $x^{SI*} < x_t^{SI*}$  with*

$$1/2 < t < 1. \quad (6)$$

**Proposition 4.** *Under a pure innovation investment strategy, the supply chain can be coordinated by promising an appropriate proportion.*

The promised proportion  $t$  can stimulate the manufacturer to join the alliance. The more the proportion undertaken by the retailer, the more operational innovation would be. A reasonable promised proportion can be made to coordinate the decentralized decisions.

**3.2. Pure Advertising Investment Strategy.** A supply chain center can then use an advertising investment strategy to stimulate market sales, which undoubtedly increases the cost of the supply chain. Under a price-sensitive market, the supply chain center optimizes advertising investment  $y^2$ :

$$\pi^{SA}(p, y) = (D - kp + \alpha y)(p - c) - y^2. \quad (7)$$

**Lemma 5.** *A pure advertising investment strategy has optimal advertising investment  $y^*$  and retail price  $p^*$  as follows:*

$$p^{SA*} = \frac{2D + 2ck - \alpha^2c}{4k - \alpha^2}, y^{SA*} = \frac{\alpha(D - ck)}{4k - \alpha^2}. \quad (8)$$

To build a supply chain alliance, the manufacturer promises to undertake part of the advertising investment. Then, the decision-making sequence is as follows: (1) The manufacturer promises a proportion of the advertising investment (i.e.,  $t$ ) and the wholesale price (i.e.,  $w$ ). (2) Then, the seller determines the retail price and the advertising investment.

The seller's decision equation

$$\pi_r^{SA}(p, y | w, t)(D - kp + \alpha y)(p - w) - (1 - t)y^2. \quad (9)$$

The manufacturer's decision equation

$$\pi_m^{SA}(w) = (D - kp + \alpha y)(w - c) - ty^2. \quad (10)$$

**Lemma 6.** *Under a pure advertising investment strategy, the supply chain alliance exists in equilibrium as follows:*

$$\begin{aligned} p_t^{SA*} &= \frac{2k(3d + ck)(-1+t)^2 + \alpha^2(ck(-1+t) + D(-1+2t))}{k(8k(-1+t)^2 + \alpha^2(-2+3t))}, \\ y_t^{SA*} &= -\frac{\alpha(D - ck)(-1+t)}{8k(-1+t)^2 + \alpha^2(-2+3t)}, \\ Q_t^{SA*} &= \frac{2k(D - ck)(-1+t)^2}{8k(-1+t)^2 + \alpha^2(-2+3t)}, \\ w_t^* &= \frac{4k(D + ck)(-1+t)^2 + \alpha^2(ck(-1+t) + D(-1+2t))}{k(8k(-1+t)^2 + \alpha^2(-2+3t))}. \end{aligned} \quad (11)$$

The seller would made the advertising investment  $y_t^{SA*}$  and then charged the retail price  $p_t^{SA*}$  to make the market sale as  $Q_t^{SA*}$  with manufacturer's promised proportion  $t$  and charged wholesale price  $w_t^*$ .

**Proposition 7.** *Under a pure advertising investment strategy, the promised proportion has a significant effect on advertising investment. Specifically,  $y_t^{SA*} \geq y^{SA*}$  with  $t \geq 1/2$ ;  $y_t^{SA*} < y^{SA*}$  with  $t < 1/2$ .*

**Proposition 8.** *Under a pure advertising investment strategy, the supply chain alliance cannot be coordinated by the promised proportion.*

Similarly, the promised proportion  $t$  can stimulate seller to join the alliance with the coordinative advertising. The more the proportion undertaken by manufacturer, the more advertising investment would be. Unfortunately, the promised proportion cannot coordinate the decentralized decisions.

**3.3. Hybrid Investment Strategy.** A supply chain alliance can make decisions simultaneously—that is, make innovation investments in the upstream and advertising investments in the downstream. Under a hybrid investment strategy, the

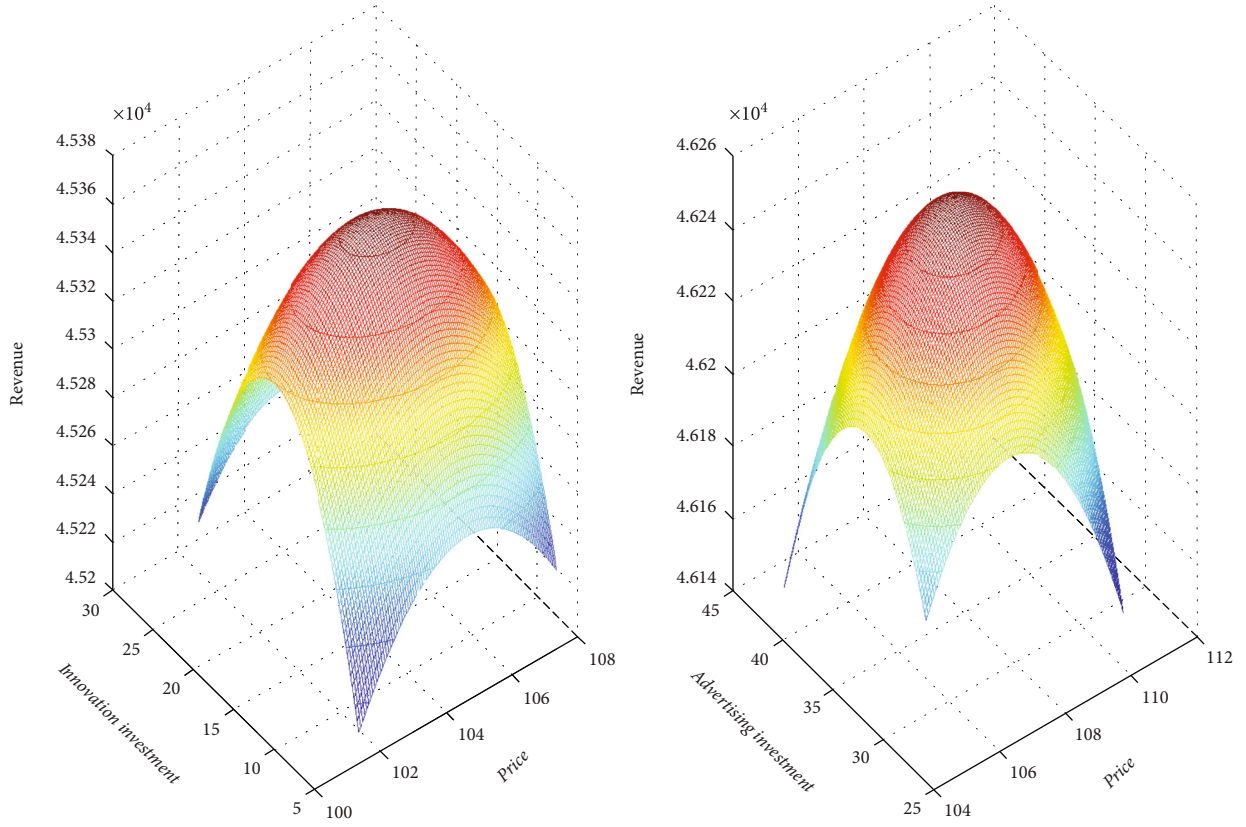


FIGURE 1: Revenue curves in pure innovation strategies (left) and pure advertising strategies (right).

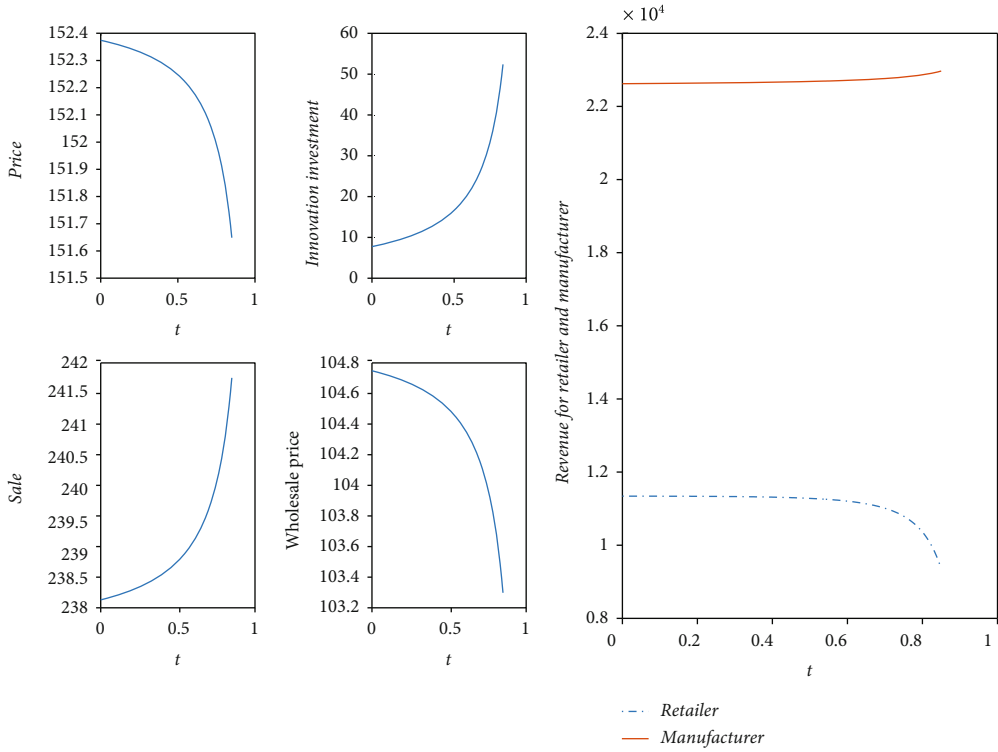


FIGURE 2: Effects of promised proportions under a pure innovation strategy.

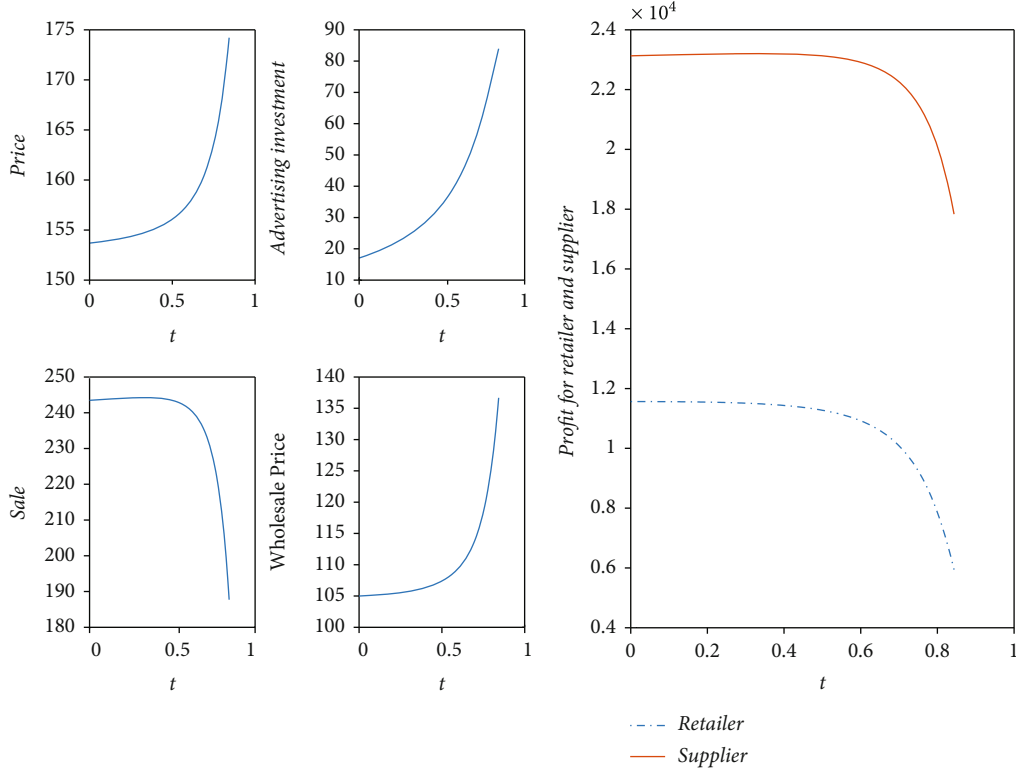


FIGURE 3: Effects of shared proportion under a pure advertising strategy.

supply chain center optimizes the equation for innovation investment and advertising investment:

$$\pi^{DC}(p, x, y) = (p - c + \beta x)(D - kp + \alpha y) - x^2 - y^2. \quad (12)$$

**Theorem 9.** Under hybrid investment strategies, the supply chain has the optimal investment and retail price, as

$$\begin{aligned} x^{DC*} &= \frac{\beta k(-D + ck)}{\alpha^2 + k(-4 + \beta^2 k)}, \\ y^{DC*} &= \frac{\alpha(-D + ck)}{\alpha^2 + k(-4 + \beta^2 k)}, \\ p^{DC*} &= \frac{\alpha^2 c - 2ck + D(-2 + \beta^2 k)}{\alpha^2 + k(-4 + \beta^2 k)}. \end{aligned} \quad (13)$$

**Proposition 10.** Under a hybrid strategy, the innovation investment is less than that in a pure innovation strategy, while the advertising investment is larger than that in a pure advertising strategy in centralized decisions.

Compared with pure strategy, the supply chain center would adjust the investment structure by decreasing innovation investments but increasing advertising investment due to the sensitivity under hybrid strategy.

Propositions 3 and 7 suggest that the initiator's promised proportion has a significant effect on the follower's investment but cannot avoid the emergence of opportunism. Hence, the supply chain alliance can make the upstream fully responsible for the innovation investment and the down-

stream fully responsible for the advertising investment. Then, the sequence of decision-making is as follows: (1) The manufacturer first promises to undertake innovation investment to improve the reproduction line, and (2) the seller sets the advertising investment and retail price according to the manufacturer's promise and the market situation. The decision functions for both parties are, respectively,

$$\pi_r^{DD}(p, y) = (D - kp + \alpha y)(p - w) - y^2, \quad (14)$$

$$\pi_s^{DD}(w, x) = (D - kp + \alpha y)(w - c + \beta x) - x^2. \quad (15)$$

**Theorem 11.** Under a hybrid strategy, the supply chain exists in equilibrium for both manufacturer and seller as

$$\begin{aligned} x^{DD*} &= \frac{\beta k(-D + ck)}{2\alpha^2 + k(-8 + \beta^2 k)}, \\ w^{DD*} &= \frac{\alpha^2(D + ck) + k(-4ck + D(-4 + \beta^2 k))}{k(2\alpha^2 + k(-8 + \beta^2 k))}, \\ y^{DD*} &= \frac{\alpha(-D + ck)}{2\alpha^2 + k(-8 + \beta^2 k)}, \\ p^{DD*} &= \frac{\alpha^2(D + ck) + k(-2ck + D(-6 + \beta^2 k))}{k(2\alpha^2 + k(-8 + \beta^2 k))}, \\ Q^{DD*} &= \frac{2k(-D + ck)}{2\alpha^2 + k(-8 + \beta^2 k)}. \end{aligned} \quad (16)$$

**Proposition 12.** Under a hybrid strategy, innovation and advertising investments are both larger than those in a

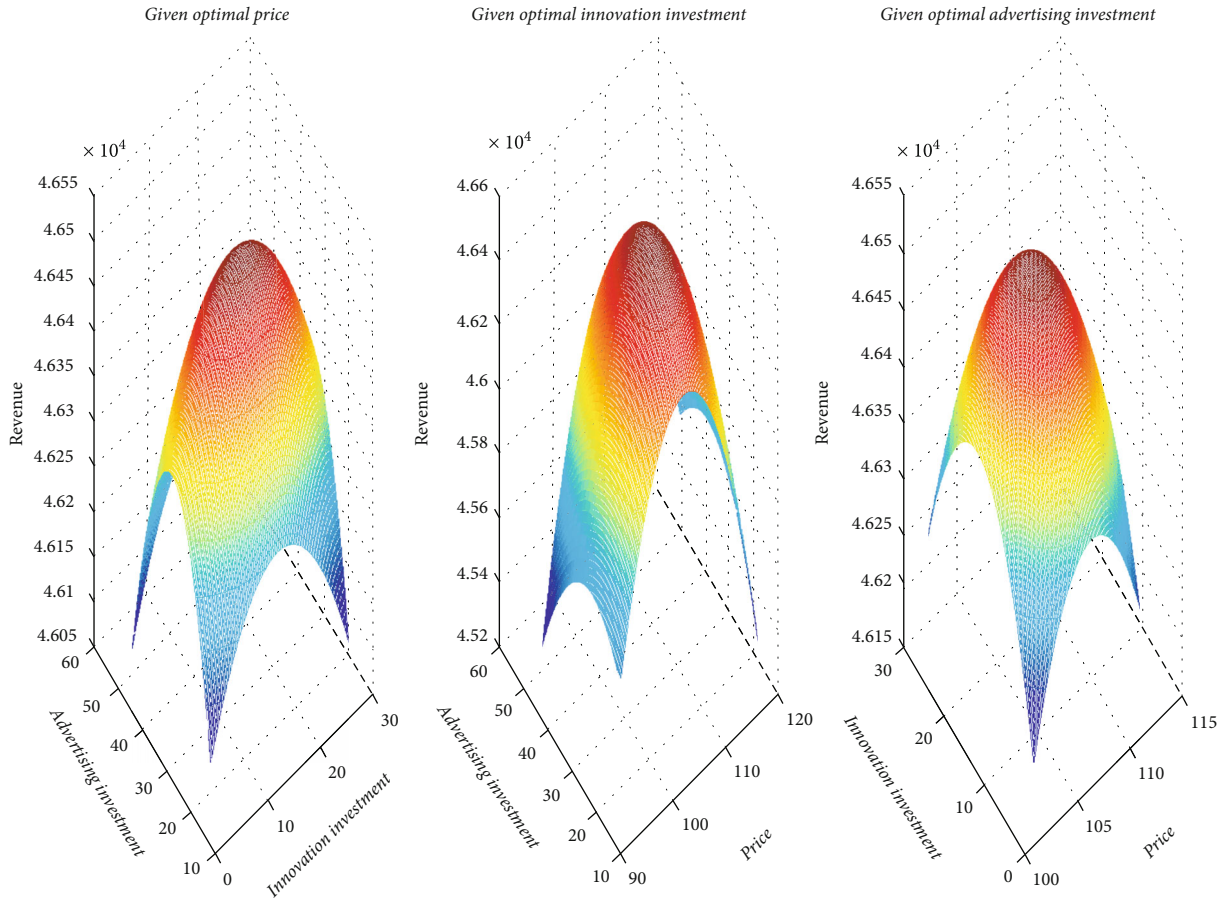


FIGURE 4: Revenue curves under a hybrid strategy given one variable.

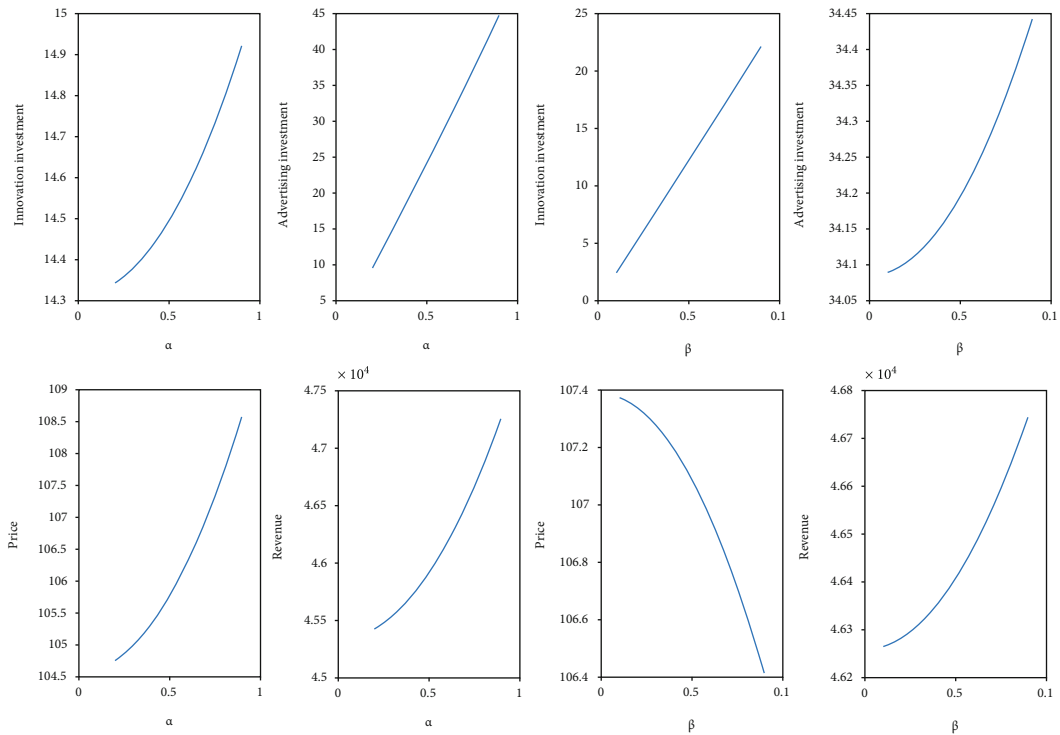


FIGURE 5: Effects of sensitive coefficients in a supply chain alliance under a hybrid strategy.

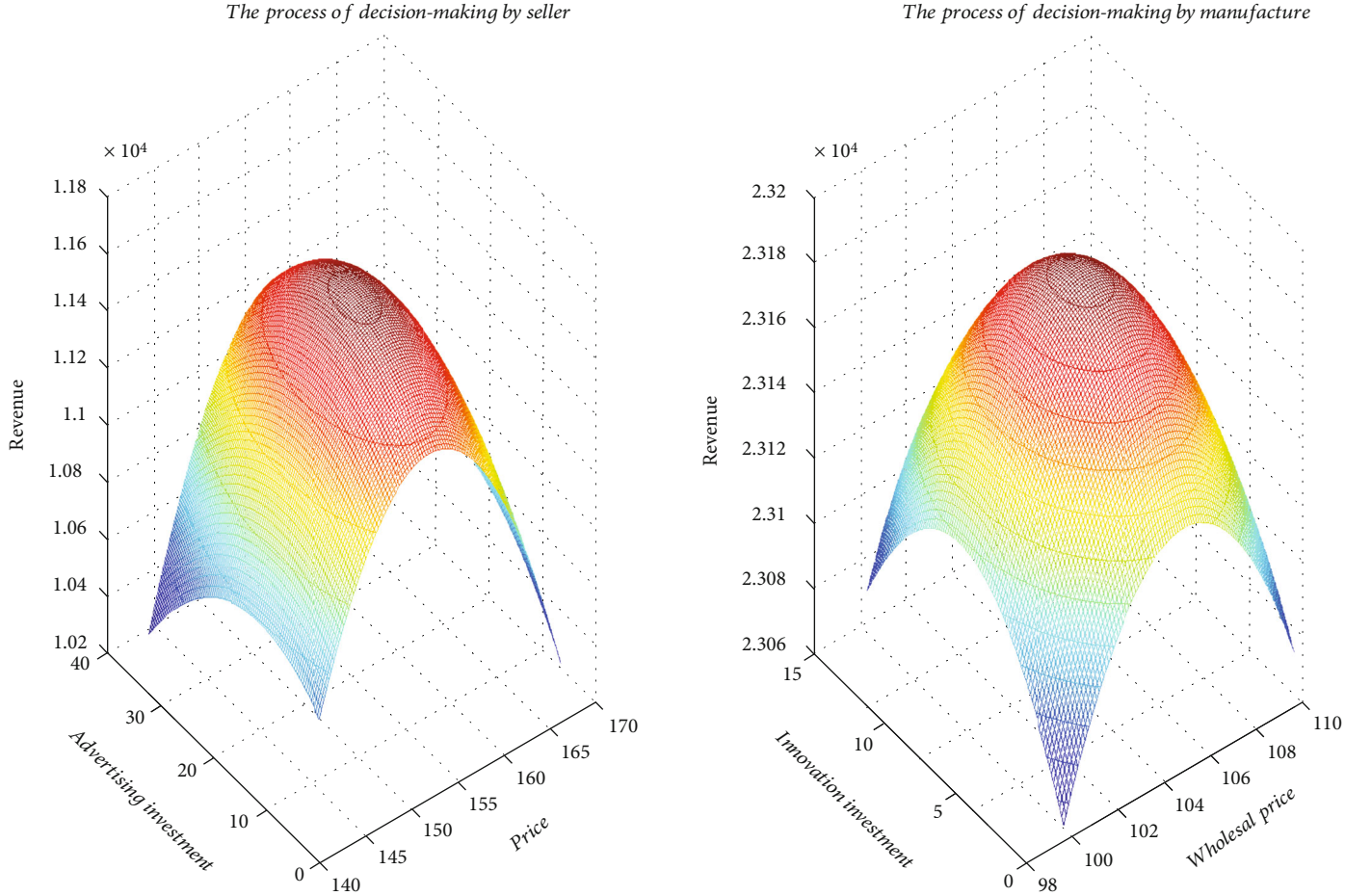


FIGURE 6: Decision-making processes of a seller and manufacturer under a hybrid strategy.

decentralized system but less than those in a pure center investment strategy. Innovation and advertising investments are both less than those in a pure strategy if the initiator's promised proportion satisfies

$$t \geq \max \left( \frac{1}{2}, \frac{\alpha^2/4k, (8k + \beta^2k^2 - \alpha^2 - A)/16k}{A} \right), \text{ and} \quad (17)$$

$$A = \sqrt{-32\beta^2k^3 + (\alpha^2 - k(8 + \beta^2k))^2}.$$

Proposition 12 suggests that actors in the supply chain alliance will reduce their investments under their full responsibility under hybrid strategy. The initiator's promised proportion can stimulate the follower to invest more investments than those in hybrid with its responsibility.

These theorems and propositions show that a supply chain alliance can be obtained but cannot be balanced in some scenarios. Although investments under various strategies can be compared analytically, it is difficult to analytically compare some decisions, such as those pertaining to profits and retail prices, in various investment strategies due to complexity. Hence, we will employ numerical analysis to demonstrate the decision-making processes.

## 4. Numerical Analysis

Before performing the numerical analysis, we need to reorganize the constraints among the variables to obtain meaningful conclusions. Statistical big data based on advanced product identification can capture the basic relationships among the variables [36–45]. Subjecting to all constraints being satisfied, it is reasonable to assume the values of some parameters as follows: potential initial market  $D=1000$ , original production cost  $c=10$ , sensitivity coefficient of market to price  $k=10$ , advertising sensitivity coefficient  $\alpha=0.7$ , and innovation sensitivity coefficient  $\beta=0.065$ .

**4.1. Pure Investment Strategies.** In pure investment strategies, the supply chain can make either innovation investments or advertising investments. The revenue curves in pure investment strategies are presented in Figure 1, as follows.

Lemmas 1 and 5 show that supply chains exist with optimal investment decisions in pure innovation and advertising strategies. Propositions 3 and 7 show that the promised proportion has a significant effect on decisions. Figure 2 presents the effects of promised proportions on decisions (e.g., retail price, innovation investment, market sales, wholesale price, revenue) in a pure innovation strategy.

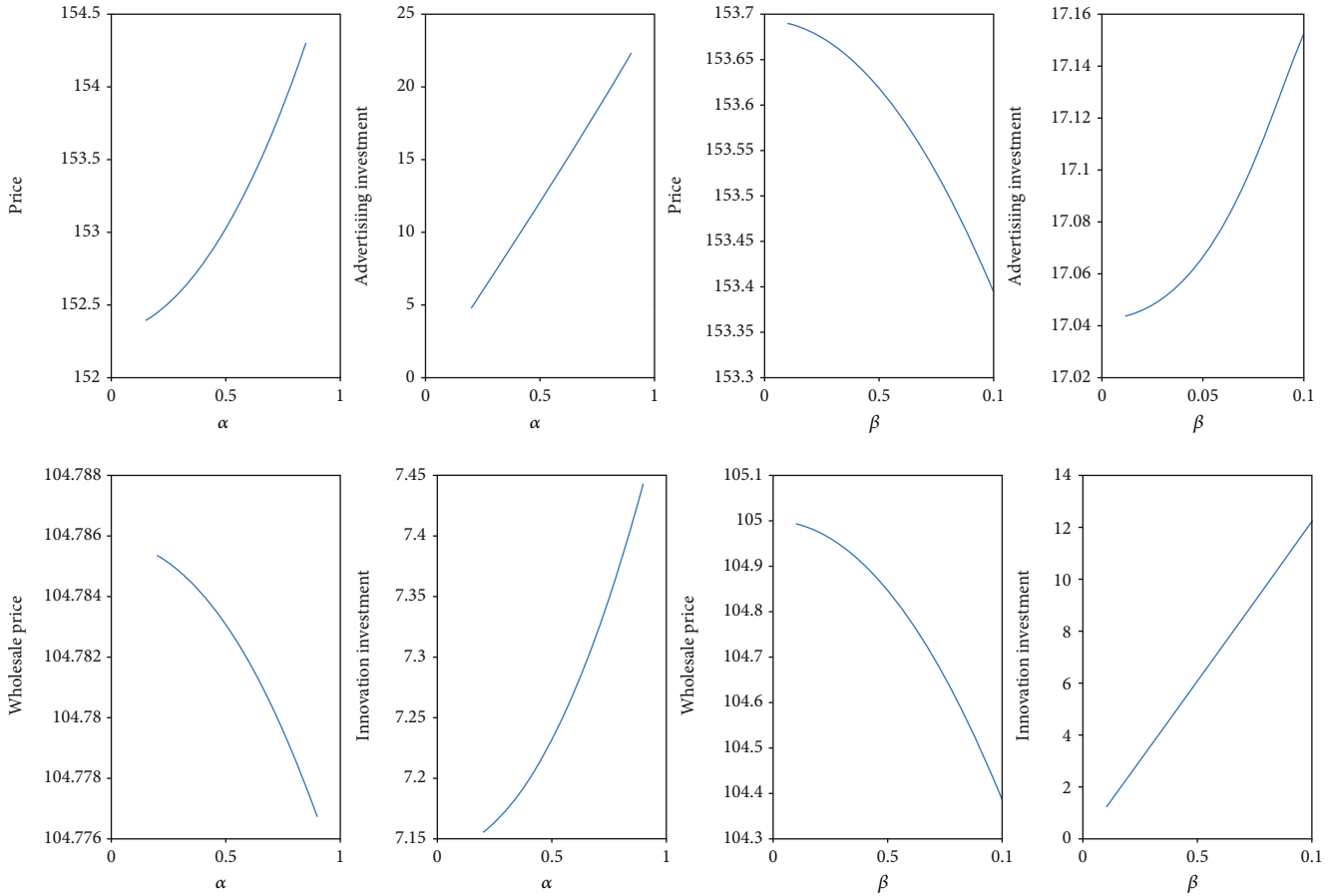


FIGURE 7: Effects of sensitive coefficients on decentralized decisions under a hybrid strategy.

The total innovation (follower made) increases with the promised proportion increase. Innovation investment reduces the reproduction cost and then decreases the wholesale price, which decrease the retail price and increases market sales. Hence, the promised proportion can stimulate the follower to complete the mission in the supply chain. The manufacturer can transform the profits from innovation investment to the seller with low wholesale prices. The seller can then decrease the retail price to enhance market competitiveness, which is the original intention of the initiator in building a supply chain alliance. However, a high promised proportion increases the manufacturer's revenue but undoubtedly increases the seller's costs and ultimately reduces the seller's revenues.

Similarly, Figure 3 shows the effects of promised proportion on decisions under a pure advertising strategy.

Advertising investment increases with the promised proportion increase. Advertising investment aims to stimulate market sales, but it also means additional costs for both the follower and the initiator. Then, the additional cost causes both the retail price and the wholesale price to increase. Then, market sales will increase for a short period and then drop rapidly. A pure advertising strategy could thus damage both actors' revenues.

**4.2. Hybrid Investment Strategies.** Figure 4 shows the revenues in the supply chain decision-making process under a hybrid strategy. Then, Figure 5 presents the effects of sensitive coefficients on optimal decisions.

Figure 5 indicates that whichever sensitive coefficient increases, innovation investment, advertising investment, and supply chain alliance revenue all increase. Intuitively, the higher the sensitivity coefficient (i.e., the higher the investment benefit), the greater the investment. This phenomenon reflects the fact that decision-making in a supply chain alliance emphasizes balance between upstream and downstream investment. This balance certainly aims to improve performance. Similar to a pure investment strategy, the advertising effect leads to a high retail price, while the innovation effect reduces the retail price via reducing the reproduction cost. However, it should be emphasized that increasing advertising investment can improve performance, which may be the main difference between the hybrid strategy and the pure strategy. The main reasons may be found in the linkage effect and balance effect in the supply chain alliance.

Figure 6 presents two players' decision-making processes in a decentralized supply chain.

To investigate the effects of sensitive coefficients on equilibrium, Figure 7 presents the optimal decisions (i.e., investment, retail price, and wholesale price changes) with sensitive coefficients (i.e.,  $\alpha$ ,  $\beta$ ).

Similarly, whichever sensitive coefficient increases, the investments made by different individual actors also crease. A linkage effect still exists in the decentralized decision-making, even though there is no cross-investment. Even in a decentralized supply chain, the manufacturer will increase

innovation when the sensitive coefficients of innovation increase and then reduce the wholesale price and finally stimulate the seller to lower the retail price. The seller also increases the advertising investment when the sensitive coefficient of advertising increases but still increases the retail price, even when the manufacturer decreases the wholesale price. The reason could be that opportunism exists in the downstream based on the advantage of being close to the market.

Promised proportion is an effective approach for building a supply chain alliance, but it has different effects on decisions under different pure investment strategies. An innovation investment strategy can transform benefits to the downstream, while an advertising investment strategy may have the opposite effect. Although these effects can be improved under a hybrid strategy, opportunism might also arise.

## 5. Conclusion

In order to cope with rigid environment regulations, manufacturing forms can initiate closed-loop supply chain alliances through suitable partners to focus on product recovery management and even to enjoy the economic benefits. However, an effective closed loop supply chain can simultaneously consider innovation investment in the upstream and advertising investment in the downstream to enjoy the economic benefits. Unlike the current closed-loop supply chain studies with one-way input decisions, this study considers the optimal investment allocation on/between the advertising and innovation within a closed-loop supply chain system, which can be initiated by an initiator's promised proportion in investment. The closed-loop supply chain alliance can be initiated by an initiator's promised proportion in investment. Two ways of building a closed-loop supply chain alliance were considered in this study: the manufacturer initiates or the retailer initiates. A manufacturer initiates a supply chain alliance by promising a proportion of the advertising investment, while the seller promises a proportion of the innovation investment. The equilibriums under various strategies were considered analytically, and some sensitivities of decisions were investigated using numerical analysis.

The results showed that under a pure investment strategy, a promised proportion has a significant effect on investment decisions and can make the supply chain achieve equilibrium. However, a reasonable promised proportion can coordinate the supply chain under a pure innovation strategy but not under a pure advertising strategy. Compared to pure investment, the supply chain center will decrease upstream investment but increase downstream investment under a hybrid strategy. Upstream investments always decrease wholesale and retail prices, while downstream investments increase retail and wholesale prices. Upstream and downstream investments will both increase with a sensitive coefficient increase. Increasing innovation investment can benefit the manufacturer but not the seller, while increasing advertising investment creates costs for both actors under a pure investment strategy. However, increasing innovation or advertising can bring benefits to both actors under a hybrid investment strategy. Generally, the upstream transforms the benefit to the downstream while downstream, not vice versa. The downstream may exhibit

opportunism based on the advantage of being close to the market. In summary, a hybrid investment strategy provides insight into optimizing the supply chain system in investments. A supply chain alliance should balance investments in the upstream and downstream simultaneously. The balance between the upstream and downstream can eliminate the negative effect of one-sided investment. Increasing innovation investment in the upstream may be a good choice for the supply chain; after all, it can transfer the benefits to the downstream.

A limitation of this study is that the models only considered operational innovation and advertising investment. There are additional investment choices that could potentially provide mutual benefit. More investment choices could be considered in future work. Moreover, the stochastic of corresponding output with investment could affect investment decisions. How the stochastic of the investment affects the investment decision could also be an interesting direction for future work.

## Data Availability

All data are in the article.

## Conflicts of Interest

The authors have declared that no competing interests exist concerning this study.

## Acknowledgments

This work was supported by the Provincial Nature Science Foundation of Guangdong (2018A030313488, 2016A060311539), Guangdong Province Young Innovative Talents Program in Universities (Natural Science) (no. 2018KQNCX327), Guangdong Province Philosophy and Social Sciences "13th Five-Year" Planning Project for Building Discipline (GD17XGL56), Guangdong Soft Science Research Project (2018A070712005), and the Fundamental Research Funds for the Central Universities (2018MSXM01).

## Supplementary Materials

Supplementary Materials: proofs of lemmas, propositions, and theorems. (*Supplementary Materials*)

## References

- [1] C. Gobbi, "Designing the reverse supply chain: the impact of the product residual value," *International Journal of Physical Distribution & Logistics Management*, vol. 41, no. 8, pp. 768–796, 2011.
- [2] C. Lu, L. Zhang, Y. Zhong et al., "An overview of e-waste management in China," *Journal of Material Cycles and Waste Management*, vol. 17, no. 1, pp. 1–12, 2015.
- [3] J. Yu, E. Williams, M. Ju, and C. Shao, "Managing e-waste in China: policies, pilot projects and alternative approaches," *Resources Conservation and Recycling*, vol. 54, no. 11, pp. 991–999, 2010.



- [4] K. Govindan, H. Soleimani, and D. Kannan, "Reverse logistics and closed-loop supply chain: a comprehensive review to explore the future," *European Journal of Operational Research*, vol. 240, no. 3, pp. 603–626, 2015.
- [5] J. Q. F. Neto, G. Walther, J. Bloemhof, J. A. E. E. van Nunen, and T. Spengler, "From closed-loop to sustainable supply chains: the WEEE case," *International Journal of Production Research*, vol. 48, no. 15, pp. 4463–4481, 2009.
- [6] Y. Zhao, Y. Cao, H. Li et al., "Bullwhip effect mitigation of green supply chain optimization in electronics industry," *Journal of Cleaner Production*, vol. 180, pp. 888–912, 2018.
- [7] D. S. Rogers and R. Tibben-Lembke, "An examination of reverse logistics practices," *Journal of Business Logistics*, vol. 22, no. 2, pp. 129–148, 2011.
- [8] M. T. Islam and N. Huda, "Reverse logistics and closed-loop supply chain of waste electrical and electronic equipment (WEEE)/E-waste: a comprehensive literature review," *Resources, Conservation and Recycling*, vol. 137, pp. 48–75, 2018.
- [9] E. Özceylan and T. Paksoy, "A mixed integer programming model for a closed-loop supply-chain network," *International Journal of Production Research*, vol. 51, no. 3, pp. 718–734, 2013.
- [10] H. Soleimani, M. Seyyed-Esfahani, and G. Kannan, "Incorporating risk measures in closed-loop supply chain network design," *International Journal of Production Research*, vol. 52, no. 6, pp. 1843–1867, 2014.
- [11] H. Krikke, "Revealing an invisible giant: A comprehensive survey into return practices within original (closed-loop) supply chains," *Resources Conservation and Recycling*, vol. 73, no. 2, pp. 239–250, 2013.
- [12] V. D. R. Guide Jr. and L. N. Van Wassenhove, "OR FOR-UM—the evolution of closed-loop supply chain research," *Operations Research*, vol. 57, no. 1, pp. 10–18, 2009.
- [13] S. Agrawal, R. K. Singh, and Q. Murtaza, "A literature review and perspectives in reverse logistics," *Resources, Conservation and Recycling*, vol. 97, pp. 76–92, 2015.
- [14] M. Liu, C. Liu, L. Xing, F. Mei, and X. Zhang, "Study on a tolerance grading allocation method under uncertainty and quality oriented for remanufactured parts," *The International Journal of Advanced Manufacturing Technology*, vol. 87, no. 5-8, pp. 1265–1272, 2016.
- [15] R. C. Savaskan, S. Bhattacharya, and L. N. V. Wassenhove, "Closed-loop supply chain models with product remanufacturing," *Management Science*, vol. 50, no. 2, pp. 239–252, 2004.
- [16] M. Fleischmann, J. M. Bloemhof-Ruwaard, R. Dekker, J. A. E. E. Nunen, and L. N. V. Wassenhove, "Quantitative models for reverse logistics: a review," *European Journal of Operational Research*, vol. 103, no. 1, pp. 1–17, 1997.
- [17] V. D. R. Guide, V. Jayaraman, R. Srivastava, and W. C. Benton, "Supply-chain Management for Recoverable Manufacturing Systems," *Interfaces*, vol. 30, no. 3, pp. 125–142, 2000.
- [18] G. C. Souza, "Closed-loop supply chains: a critical review, and future research," *Decision Sciences*, vol. 44, no. 1, pp. 7–38, 2013.
- [19] D. Stindt and R. Sahamie, "Review of research on closed loop supply chain management in the process industry," *Flexible Services and Manufacturing Journal*, vol. 26, no. 1-2, pp. 268–293, 2014.
- [20] K. Schweiger and R. Sahamie, "A hybrid Tabu search approach for the design of a paper recycling network," *Transportation Research Part E: Logistics and Transportation Review*, vol. 50, no. 1, pp. 98–119, 2013.
- [21] X. Zhu and X. Xu, "An integrated optimization model of a closed-loop supply chain under uncertainty," in *LISS 2012*, Z. Zhang, R. Zhang, and J. Zhang, Eds., Springer, Berlin, Heidelberg, 2013.
- [22] S. Mitra, "Periodic review policy for a two-echelon closed-loop inventory system with correlations between demands and returns," *Opsearch*, vol. 50, no. 4, pp. 604–615, 2013.
- [23] A. El korchi and D. Millet, "Designing a sustainable reverse logistics channel: the 18 generic structures framework," *Journal of Cleaner Production*, vol. 19, no. 6-7, pp. 588–597, 2011.
- [24] R. C. Savaskan and L. N. V. Wassenhove, "Reverse channel design: the case of competing retailers," *Management Science*, vol. 52, no. 1, pp. 1–14, 2006.
- [25] T. M. Choi, Y. Li, and L. Xu, "Channel leadership, performance and coordination in closed loop supply chains," *International Journal of Production Economics*, vol. 146, no. 1, pp. 371–380, 2013.
- [26] R. Subramanian, M. E. Ferguson, and L. B. Toktay, "Remanufacturing and the component commonality decision," *Production and Operations Management*, vol. 22, no. 1, pp. 36–53, 2013.
- [27] B. C. Giri and S. Sharma, "Manufacturer's pricing strategy in a two-level supply chain with competing retailers and advertising cost dependent demand," *Economic Modelling*, vol. 38, no. 38, pp. 102–111, 2014.
- [28] Y.-W. Zhou, J. Li, and Y. Zhong, "Cooperative advertising and ordering policies in a two-echelon supply chain with risk-averse agents," *Omega*, vol. 75, no. C, pp. 97–117, 2018.
- [29] J. L. Hartley, B. J. Zirger, and R. R. Kamath, "Managing the buyer-supplier interface for on-time performance in product development," *Journal of Operations Management*, vol. 15, no. 1, pp. 57–70, 1997.
- [30] B. Kim, "Coordinating an innovation in supply chain management," *European Journal of Operational Research*, vol. 123, no. 3, pp. 568–584, 2000.
- [31] F. T. S. Chan, A. Nayak, R. Raj, Y. L. Chong, and T. Manoj, "An innovative supply chain performance measurement system incorporating Research and Development (R&D) and marketing policy," *Computers and Industrial Engineering*, vol. 69, no. 3, pp. 64–70, 2014.
- [32] S. M. Gilbert and V. Cvsa, "Strategic commitment to price to stimulate downstream innovation in a supply chain," *European Journal of Operational Research*, vol. 150, no. 3, pp. 617–639, 2003.
- [33] W. Xu, S. Q. Zheng, D. P. Chen, and Nanjing Audit University, "Study on collaboration and innovation model of manufacturers&sellers in the closed-loop supply chain," *Chinese Journal of Management Science*, vol. 22, no. 7, pp. 116–123, 2014.
- [34] V. Agrawal and S. Seshadri, "Impact of uncertainty and risk aversion on Price and order quantity in the newsvendor problem," *Manufacturing & Service Operations Management*, vol. 2, no. 4, pp. 410–423, 2000.
- [35] J. Xie and J. C. Wei, "Coordinating advertising and pricing in a manufacturer-retailer channel," *European Journal of Operational Research*, vol. 197, no. 2, pp. 785–791, 2009.
- [36] L. Fabisiak, "Web service usability analysis based on User preferences," *Journal of Organizational and End User Computing*, vol. 30, no. 4, pp. 1–13, 2018.

- [37] J. Jakli, T. Grubljei, and P. S. Coelho, "The shift to socio-organizational drivers of business intelligence and analytics acceptance," *Journal of Organizational and End User Computing*, vol. 31, no. 2, pp. 37–64, 2017.
- [38] Y. Jiang, X. Gu, D. Wu et al., "A novel negative-transfer-resistant fuzzy clustering model with a shared cross-domain transfer latent space and its application to brain CT image segmentation," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 99, pp. 1–13, 2019.
- [39] Y. Jiang, Y. Zhang, C. Lin, D. Wu, and C. T. Lin, "EEG-based driver drowsiness estimation using an online multi-view and transfer TSK fuzzy system," *IEEE Transactions on Intelligent Transportation Systems*, vol. 99, pp. 1–13, 2020.
- [40] G. Khatwani and P. R. Srivastava, "Impact of information technology on information Search Channel selection for consumers," *Journal of Organizational and End User Computing*, vol. 30, no. 3, pp. 63–80, 2018.
- [41] A. Shahri, M. Hosseini, K. Phalp, J. Taylor, and R. Ali, "How to engineer Gamification," *Journal of Organizational and End User Computing*, vol. 31, no. 1, pp. 39–60, 2019.
- [42] N. Suyel and R. Pinki, "PpBAC: popularity based access control model for cloud computing," *Journal of Organizational and End User Computing*, vol. 30, no. 4, pp. 14–31, 2018.
- [43] Y. Zhang, F. L. Chung, and S. Wang, "Clustering by transmission learning from data density to label manifold with statistical diffusion," *Knowledge-Based Systems*, vol. 193, article 105330, 2019.
- [44] Y. Zhang, J. Li, X. Zhou, T. Zhou, and J. Yang, "A view-reduction based multi-view TSK fuzzy system and its application for textile color classification," *Journal of Ambient Intelligence and Humanized Computing*, vol. 1, pp. 1–11, 2019.
- [45] Y. Zhang, S. Wang, K. Xia, Y. Jiang, and P. Qian, "Alzheimer's disease multiclass diagnosis via multimodal neuroimaging embedding feature selection and fusion," *Information Fusion*, vol. 66, pp. 170–183, 2021.

## Research Article

# Markdown Time for Perishables Based on Dynamic Quality Evaluation for Complex Data Analysis

Jiang-Tao Wang,<sup>1</sup> Jian-Jun Yu,<sup>2</sup> Yu-Hsi Yuan ,<sup>3</sup> Sang-Bing Tsai ,<sup>4</sup> and Shu-Fen Zhang <sup>5</sup>

<sup>1</sup>Zhongshan Institute, University of Electronic Science and Technology of China, Zhongshan 528400, China

<sup>2</sup>School of Business Administration, South China University of Technology, Guangzhou 510641, China

<sup>3</sup>Department of Labor & Human Resources, Chinese Culture University, Taipei 111, Taiwan

<sup>4</sup>Regional Green Economy Development Research Center, School of Business, WUYI University, Wuyishan 354300, China

<sup>5</sup>School of Chinese Medicine, Guangdong Pharmaceutical University, Guangzhou 510006, China

Correspondence should be addressed to Yu-Hsi Yuan; [yuanyh@gm.ypu.edu.tw](mailto:yuanyh@gm.ypu.edu.tw) and Shu-Fen Zhang; [zsf063@gdpu.edu.cn](mailto:zsf063@gdpu.edu.cn)

Received 9 December 2020; Revised 28 December 2020; Accepted 31 December 2020; Published 3 February 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Jiang-Tao Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Perishables' freshness is a dynamic state, which can be captured by statistical big data based on advanced product identification and sensory technologies. A time-based markdown strategy model is proposed to consider the markdown strategy in maximizing performance. The results showed that demand uncertainty would force them to implement a markdown strategy in advance, merchants' profit and initial inventory are sensitive to the uncertainty and deterioration rate of perishable quality, and markdown time has a regulatory effect on initial inventory.

## 1. Introduction

Freshness is becoming a quality criterion of great importance to consumers regarding perishable products (Peneau et al. [1]). The freshness as well as the price has been viewed as an important factor influencing consumers' purchase decision. Many companies are making efforts to promote the fresh concept in the retail market. Perishable operation management must consider how to deal with perishables before they spoil to ensure food safety. Sales at discounted prices have become a common strategy in retail processes, especially with long lead times and hard-to-predict demand. As a common method of matching supply and demand, the markdown strategy has a dramatic effect on a retailer's profitability [2], i.e., too early or too deeply will lead to lost revenue, while delaying markdowns or keeping them shallow will lead to liquidating inventory at even lower prices at the end of the season [3]. Hence, effective markdown strategies should not ignore the timing of the markdown as well as the initial inventory.

Perishable's typical characteristic is that the freshness (i.e., quality) is a dynamic state that decreases continuously to the point when it is unfit for consumption. Most advanced product identification and sensory technologies such as radio frequency identification technology (RFID) and time temperature indicator (TTI) have been developed to make it possible to capture the perishables' quality information, such as product identity, properties, and related data [4–8]. These rich quality index data can be used to predict dynamically the shelf life. At present, some research literature of perishable products focus on the preservation technology investment [9–11] and some research considered the price and inventory problem of perishables by considering the markdown time as an exogenous variable [12–14]. However, how to effectively implement markdown strategies consisting of time and inventory based on the dynamic shelf life information should be an important issue in perishable operation management, especially in an uncertain demand environment.

A time-sensitive markdown strategy is employed to consider the markdown time challenge based on the time-variance market sale model. This time-variance markdown model can capture the dynamic quality evolution characteristics. The relations between the markdown time and the initial inventory are considered under uncertainty market demand. Partial decision-making attributes are demonstrated through numerical analysis. The results suggest that market uncertainty may not always be detrimental to merchants and may even increase the ultimate benefit to them. Both the merchant's profit and initial inventory are sensitive to the market demand uncertainty while the markdown time is not.

The rest of this paper is organized as follows. Section 2 presents a review of the existing literature related to our research. The model description is presented in Section 3. We will consider the markdown strategy under deterministic and uncertain market demand, respectively, in Sections 4 and 5. Section 6 presents numerical results based on the theoretical results. A summary of this research and discussion of future research directions are conducted in Section 7. Some proofs are presented in supplementary files (available here).

## 2. Literature Review

Perishables' production and inventory control problems have been given more attention in the inventory literature [12, 15]. In the study of perishables, the research relating to this study can be grouped into two streams.

A stream of related research literature focuses on the preservation technology investment. These researches consider how to control and reduce the deterioration of perishables through effective procedural changes and specialized equipment acquisition. The main theme is to decide the investment in preservation technologies to reduce deterioration rates using product inventory models [9–11]. For example, Shah et al. [16] derived the optimal preservation technology investment, retail price, and purchase quantity for inventory systems where units are subject to constant rates of deterioration.

Another stream of related research literature focuses on the price and inventory control model, which either considers the product with fixed lifetimes, random lifetimes, and decay rates that correspond to proportional inventory decreases in terms of utility or quantity [14, 17, 18]; or focuses on combining pricing and inventory controls to investigate how different pricing modalities affect pricing and inventory decisions [19–22]; or considers how to adjust prices to attract customers over one or more periods in a certain demand market [13, 23]. These researches suggested that customer valuations have significant effects on retailers' price and inventory decisions [24–26], and markdown strategy in operation management is a controversial topic. In particular, the markdowns would not be viewed as an optimal pricing modality for monopoly firms when the firm has high levels of inventory [27], and has often been characterized as a wasteful practice [28]. However, markdowns have also been suggested to be a more profitable strategy than single fixed prices [29] and a better pricing modality when customer behavior is considered [30, 31]. We noticed that these

researches always viewed the markdown time as an exogenous variable when considering price and/or inventory for perishables. However, this paper differs in that it views the markdown time as a decision variable and considers it under uncertainty market demand.

Demand uncertainty is an important issue in operation management literature. Perishables' market demand is also uncertain and would be influenced by the dynamic quality evolution. The newsvendor model is always used to consider the decision problem under uncertain demand [32, 33]. However, most of the extant newsvendor model literature assumed that demand uncertainty is exogenous. The time-variant market demand model in a quick response system assumed that lead-time reduction can help to enhance forecast accuracy of uncertain market demand [34–36] and was then employed to consider the lead-time decisions (purchase timing, guaranteed delivery time, lead time, and options contracts) in the supply chain [36–39]. Inspired by this stream of research, we assumed the perishables' demand uncertainty is time-sensitive and relates to the sale period length, which is also common in reality, i.e., the longer the sale period, the greater the uncertainty. The literature that is closely related to this study is [3] and that by Wang et al. [40]. However, Şen [3] considered the markdown time for perishable and substitutable produce under a determined market, and Wang et al. [40] consider the introduction time for a holiday product with similar cumulative market demand model. The main difference from the above literature is that this study assumed that the market demand rate is price- and time-sensitive, then considers the markdown time as well as the initial inventory strategy under uncertain markets.

## 3. Model Description

All perishables are available for a limited sale period and must be sold before the "sell-by-date," which has been printed on product labels by the producer to assure the product is of satisfactory quality. The perishables' freshness then can be represented in terms of remaining shelf life. The effects of remaining shelf life on freshness with the market demand are expressed commonly by the kinetic model approach in the literature [41]. An exponential quality-decay weighting function (i.e.,  $\beta q e^{-\lambda t}$ ) derived from the kinetic model approach has been widely adopted in perishables' product research [14, 42] then a price and quality sensitive expected market demand rate for perishable product can be viewed as  $f(D(t)) = f(p(t), t) = d - \alpha p(t) + \beta q(t) = d - \alpha p(t) + \beta q e^{-\lambda t}$ , where  $d(>0)$  is the maximum potential market demand rate,  $\lambda$  represents the quality deterioration rate, and  $\alpha, \beta(\geq 0)$  represent demand sensitivity to the product price and product quality, respectively. The market demand rate satisfies  $df(D(t))/dp \leq 0$ ,  $df(D(t))/dt \leq 0$ . The expected market demand during the selling season  $[0, T]$  is then  $D(T) = \int_0^T f(p(t), t) dt$ .

Merchants prepare inventory in advance and then appropriately use markdown pricing during the season to reverse the declining demand. Theoretically, dynamic pricing strategy  $p(t)$  can maintain the expected demand rate at a high

level as the quality deteriorates. However, dynamic pricing methods incur change cost, and too many markdowns represent failure in some area [43]. A single price change may be as effective as the more flexible pricing ([44]; Sen, 2013). Hence, it is reasonable to assume that merchants would have only one chance to implement the markdown strategy during the selling season. If the merchant makes the markdown price at time  $z \in (0, T)$ , then the retail price would be expressed as  $p(t) = \{p, 0 \leq t \leq z; p\theta, z < t \leq T, \theta \in (0, 1]\}$  with a price discount  $\theta$ , and then the expected market sales are divided into two parts as  $E(D_1(p(t), t)) = \int_0^z f(p(t), t) dt = \mu_{0z}(p(z), z)$ ,  $E(D_2(p(t), t)) = \int_z^T f(p(t), t) dt = \mu_{zT}(p(z), z)$ .

The errors of the actual market sales relate to the sale time [35, 40]. So, market sale can be expressed as  $D_1(p(t), t) = \mu_{0z} + \varepsilon(z)$ ,  $D_2(p(t), t) = \mu_{zT} + \varepsilon(T - z)$ , where  $\varepsilon(t)$  is the stochastic part of market sales and satisfies  $d\text{Var}(\varepsilon(t))/dt \leq 0$  (Chiu, Choi, and Li 2009). Following Chen and Chuang [35] and Wang et al. [40], we assume  $\varepsilon(t) = ((T - t)/T)\sigma X$ , where  $X \sim N(0, 1)$  with the pdf  $f(x)$  and cdf  $F(x)$  and its inverse function  $F^{-1}(x)$ . A larger  $\sigma$  means higher market demand uncertainty.  $ED(p(t), t)$  must be larger than  $\sigma$  so that  $\Pr(D(p(t), t) \leq 0)$  can be negligible.

Merchant's markdown strategies are to consider the initial inventory ( $Q$ ) and markdown time  $z$  to improve the operation performance with the quality evolving over time. In order to focus on the initial inventory and the markdown time decisions under uncertain market demand, the initial price  $p$  and the markdown price are exogenous and predetermined as assumed in [3]. Then, the merchant's optimal decision function is

$$\pi(p(t), z) = \sum_{i=1}^2 p(t) \min(Q, D_i(t)) - cQ. \quad (1)$$

#### 4. The Markdown Strategy under Deterministic Market Demand

As a benchmark, we first consider markdown strategy under the deterministic market demand. Then, the actual market demand is  $ED(p, \theta, z) = \int_0^z d - ap + \beta q e^{-\lambda t} dt + \int_z^T d - ap\theta + \beta q e^{-\lambda t} dt = (\beta q/\lambda)(1 - e^{-\lambda T}) + dT - ap(\theta T + (1 - \theta)z)$  with the predetermined price  $p(t) = \{p, 0 \leq t \leq z; p\theta, z < t \leq T, \theta \in (0, 1]\}$ , and the profit function can be viewed as  $\pi(z, Q) = p \int_0^z d - ap + \beta q e^{-\lambda t} dt + p\theta \int_z^T d - ap\theta + \beta q e^{-\lambda t} dt - cQ$  with  $Q = E(D(p, \theta, z))$ .

**Theorem 1.** *Under deterministic market demand, there exists unique optimal markdown time for the merchant's expected profit function as  $z^* = (1/\lambda) \ln(\beta q/(a(p(1 + \theta) - c) - d))$  and then the initial inventory is  $Q = (\beta q/\lambda)(1 - e^{-\lambda T}) + dT - ap(\theta T + (1 - \theta)z^*)$ .*

*Proof.* Direct derivation of the profit function is sufficient.

**Proposition 2.** *The higher the initial retail price, the earlier the markdown time; the higher the discount price, the later*

*the markdown time; and the later the markdown time, the smaller the inventory would be.*

*Proof.* Direct derivation is sufficient as  $\partial z^*/\partial p \leq 0$ ,  $\partial z^*/\partial \theta \leq 0$ .

Theorem 1 and Proposition 2 suggest that if a merchant sets a high initial price, then he will implement the price reduction early. However, if the merchant plans to take a significant price cut, he would delay the price reduction time. Eventually, the willingness to delay price reduction would lead to conservative initial inventory.

#### 5. The Markdown Strategies under Uncertain Market Demand

Under an uncertain market demand, the merchant can adopt the markdown strategy under two scenarios as follows.

**5.1. Inventory Clearance Markdown.** If the markdown price is aimed at clearing the inventory, then it can be viewed as the salvage price in most operation literature, i.e.,  $p\theta \leq c$ . Then, the merchants' profit function is  $\pi(Q, z) = E(p \min(Q, D) + p\theta(Q - D)^+) - cQ = p \min(\mu_{0z}(p(z)) + \varepsilon(z), Q) + p\theta(Q - \mu_{0z}(p(z)) - \varepsilon(z))^+ - cQ = (p - c)Q - p(1 - \theta)(T - z/T)\sigma \int_{-\infty}^{((Q - \mu_{0z})/\sigma)(T/(T - z))} F(x) dx$ .

**Lemma 3.** *For any markdown time  $z$ , the merchant's profit function is a concave in the initial inventory.*

*Proof.* Derive the profit function twice on  $Q$  and then obtain that  $\partial \pi(Q, t)/\partial Q = (p - c) - p(1 - \theta) \cdot F(((Q - \mu)/\sigma)(T/(T - z)))$ ,  $\partial^2 \pi(Q, t)/\partial Q^2 = -(p(1 - \theta)/\sigma)(T/(T - z)) \cdot f(((Q - \mu)/\sigma)(T/(T - z))) \leq 0$ .

**Theorem 4.** *To clear the inventory with the markdown price, the merchant would delay the markdown time to the expiration rate or time as the demand rate decreases to zero, i.e.,  $z^* = T$  or  $z^* = (1/\lambda) \ln(\beta q/(ap - d))$ , and the inventory is  $Q^* = (d - ap)z^* + (1 - e^{-\lambda z^*})\beta q/\lambda + \sigma(T - z^*)/T \cdot F^{-1}((p - c)/p(1 - \theta))$ .*

**Proposition 5.** *If the markdown price is aimed at clearing the inventory, the markdown time may be earlier than that under the deterministic market demand.*

*Proof.* Compare the markdown times as follows,  $a(p(1 + \theta) - c) = ap + a(p\theta - c) \leq ap$  ( $p\theta \leq c$ ), then  $(1/\lambda) \ln(\beta q/(ap - d)) \leq (1/\lambda) \ln(\beta q/(a(p(1 + \theta) - c) - d)) \leq T$ .

**5.2. Promotional Markdown.** If the markdown price is not to clear the leftover inventory at once but to reverse the declining demand caused by quality deterioration, it would occur whenever the demand rate cannot reach the expected value. Then, the markdown price would not surely clear the inventory. The market sales before and after the markdown price can be viewed as  $\min(Q, D_1)$  and  $\min(Q - \min(Q, D_1))$ ,

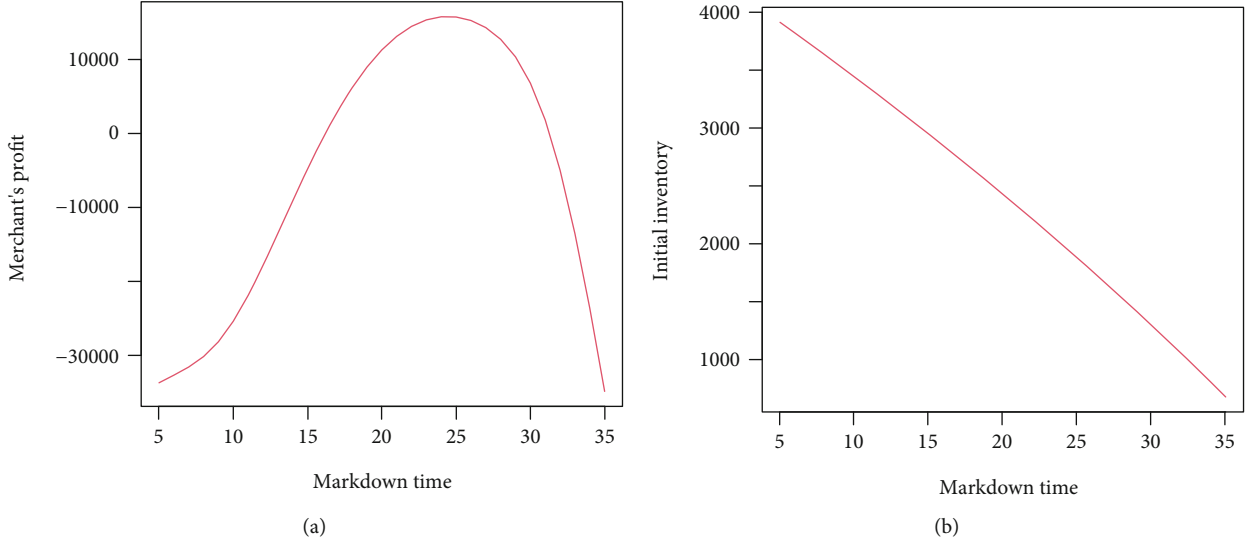


FIGURE 1: Merchant's profit (a) and initial inventory (b) on the markdown time.

$D_2$ ). The merchant's profit function can be expressed as  $\pi(Q, p) = \mathbf{E}(p \min(Q, D_1) + p\theta \min(Q - \min(Q, D_1), D_2)) - cQ$ . Drawing on the approach used by Netessine and Shumsky (2005) and noting that  $\min(Q, D_1) = \{Q, (1/\sigma)(T/(T-z))(Q - \mu_{0z}) \leq x; D_1, (1/\sigma)(T/(T-z))(Q - \mu_{0z}) \geq x\}$  and  $\min(Q - \min(Q, D_1), D_2) = \{0, (1/\sigma)(T/(T-z))(Q - \mu_{0z}) \leq x; Q - D_1, (Q - \mu_{0T})/\sigma \leq x \leq (1/\sigma)(T/(T-z))(Q - \mu_{0z}); D_2, x \leq (Q - \mu_{0T})/\sigma\}$ , the merchant's profit function can be rewritten as  $\pi(Q, p) = p \left( \int_{-\infty}^{(T/z)((Q-\mu_1)/\sigma)} (\mu_{0z} + (z/T)\sigma x) f(x) dx + \int_{(T/z)((Q-\mu_{0z})/\sigma)}^{\infty} Q f(x) dx + p\theta \left( \int_{(Q-\mu_{0z}+\mu_{zT})/\sigma}^{(T/z)((Q-\mu_{0z})/\sigma)} (Q - \mu_{0z} - (z/T)\sigma x) f(x) dx + \int_{-\infty}^{(Q-\mu_{0z}+\mu_{zT})/\sigma} (\mu_{zT} + ((T-z)/T)\sigma x) f(x) dx \right) - cQ \right) - p(1-\theta)(z/T)\sigma \int_{-\infty}^{(T/z)((Q-\mu_{0z})/\sigma)} F(x) dx - p\theta\sigma \int_{-\infty}^{(Q-\mu_{0z}+\mu_{zT})/\sigma} F(x) dx$ .

**Lemma 6.** For any markdown time  $z$ , the merchant's profit function  $\pi(Q, z)$  is a concave in the initial inventory  $Q$ . The optimal initial inventory  $Q^*(z)$  satisfies the equation as  $p - c - p(1-\theta)F((T/z)((Q - \mu_{0z})/\sigma)) - p\theta F((Q - (\mu_{0z} + \mu_{zT}))/\sigma) = 0$ .

**Lemma 7.** The initial inventory satisfies  $(z/T)\sigma F^{-1}((p-c)/p) + \mu_{0z} \leq Q^* \leq \sigma F^{-1}((p-c)/p) + \mu_{zT} + \mu_{0z}$ , i.e.,

- (1) if  $(c/p) \leq 0.5$ , then  $F^{-1}((p-c)/p) \geq 0$ , and  $0 \leq (z/T)\sigma F^{-1}((p-c)/p) + \mu_{0z} \leq Q^* \leq \sigma F^{-1}((p-c)/p) + \mu_{zT}$
- (2) if  $(c/p) \geq 0.5$ , then  $F^{-1}((p-c)/p) \leq 0$ , then  $\sigma(\geq 0)$  satisfies that

- (a) if  $\sigma F^{-1}((p-c)/p) + \mu_{zT} \geq 0$ , then  $\mu_{0z} \leq Q^* \leq \sigma F^{-1}((p-c)/p) + \mu_{0z} + \mu_{zT}$
- (b) if  $\sigma F^{-1}((p-c)/p) + \mu_{zT} \leq 0$ , then  $Q^* \leq \mu_{0z}$

**Theorem 8.** There exists an optimal markdown time  $z^*$ , which can be found by the following algorithm:

(1) For any given  $z$ , solving the equation  $p - c - p(1-\theta)F((T/(T-z))((Q - \mu_{0z})/\sigma)) - p\theta F((Q - \mu_{0T})/\sigma) = 0$  to obtain the optimal inventory  $Q^*$ ; (2) substituting the  $Q^*$  into the profit function and then obtaining the  $\pi(Q^*, z)$ ; (3) computing value  $\pi(Q^*, z)$  on  $z \in (0, T)$  and then obtaining the optimal  $z^* = \arg \max \pi(Q^*, z)$ .

*Proof.* Substituting the  $Q^*(z)$  into the merchant's profit function and then obtaining  $\pi(Q^*(z), z)$ . Note that  $\pi(Q^*(z), z)$  is a nonlinear univariate continuous function and satisfies  $\lim_{z \rightarrow 0^+} \pi(Q^*(z), z) = \pi(Q^*(0), 0)$  and  $\lim_{z \rightarrow T^-} \pi(Q^*(z), z) = (p-c)Q - p\theta\sigma \int_{-\infty}^{(Q-\mu_{0-T})/\sigma} F(x) dx$ . Actually, if  $z \rightarrow 0^+$ , the markdown strategy degenerates into a daily low price strategy. So, there must exist an optimal time  $z^* \in [0, T]$  that satisfies  $z^* = \arg \max \pi(Q^*(z), z)$ .

Lemma 7 and Theorem 8 suggest that there exists an optimal markdown time for a single price change strategy. The initial inventory is influenced by the initial price and the uncertainty in market sales. Tiny margins and large uncertainty would seriously dampen the merchant's initial inventory. Markdown strategy with time decisions can provide different insights for perishable operation management. Although the optimal initial inventory and markdown time can be derived, more decisions and the relations among the decision variables cannot be obtained analytically due to the nonlinearity of expressions. Therefore, more market insights implied in these complex relationship structures need to be explored by numerical analysis.

## 6. Numerical Analysis

Before performing the numerical analysis, we need to reorganize the constraints among parametric variables such as that the  $ED(p(t), t)$  must be larger than  $\sigma$  so that  $\Pr(D(p(t), t)$

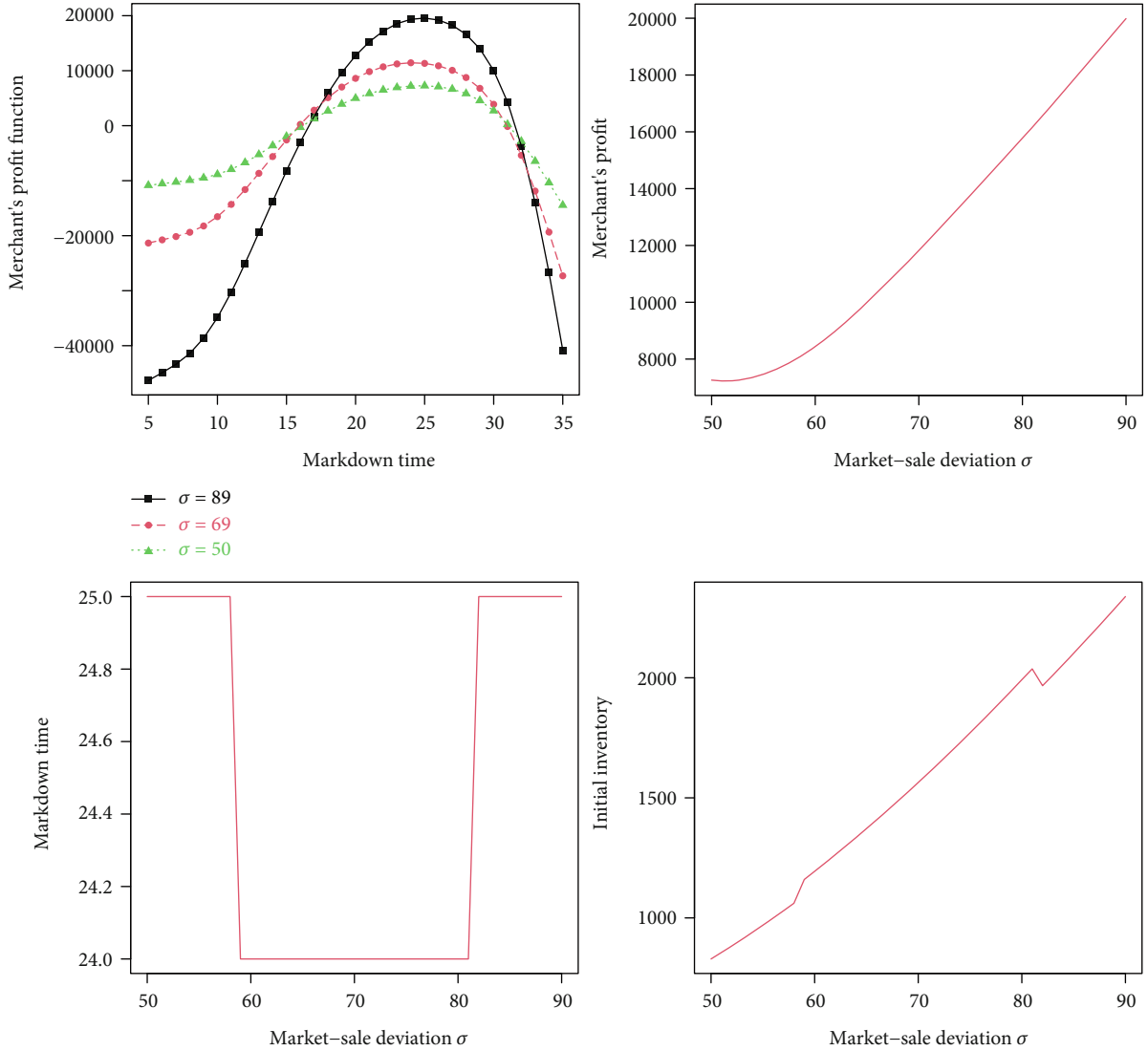


FIGURE 2: The market-sale deviation on the decisions.

$\leq 0$  can be negligible, and the parameters  $\alpha, \beta$  must satisfy that  $0 \leq ap - d \leq a(p(t)(1 + \theta) - c) - d \leq \beta q e^{-\lambda t}, t \in (0, T)$ . It is reasonable to assume the values of some parameters as market size  $d = 70$ ; initial quality  $q = 10$ ; demand sensitivity to price and quality  $a = 1$  and  $\beta = 1.5$ , respectively; quality deterioration  $\lambda = 0.02$ ; selling period length  $T = 40$ ; and product cost  $c = 22$ , initial price  $p = 80$ , and  $\theta = 0.25$ ; the stochastic part of market sales is assumed to follow the normal distribution, i.e.,  $x \in N(0, 80)$ .

**6.1. The Revenue and Initial Inventory on Markdown Time.** In order to demonstrate the effectiveness and feasibility of the algorithm in Theorem 8, Figure 1 presents the initial inventory and merchant's revenue with the markdown time evolution. Consistent with the theoretical analysis, the merchant's revenue function is a concave function on the markdown time. Hence, there exists an optimal markdown time. Besides, although the optimal initial inventory satisfies a special equation, the relationship between the initial inventory and markdown time can be approximately linear.

**6.2. The Effect of Uncertainty on the Markdown Strategies.** Based on the algorithm in Theorem 8, the merchant's profit function, final profit, markdown time, and initial inventory are presented in Figure 2. Figure 2 suggests that the merchant's profit as well as the initial inventory increases with the market-sale deviation increase. Theorem 1 points out that the merchant would delay the markdown time under a deterministic demand, which will result in small initial inventory and finally reduce the sale volumes as well as the profit during the sale period. However, uncertainty would prompt merchants to implement early price reduction, which ultimately increases the potential market sale volumes during the sale season. The markdown time does not seem to be very sensitive to the deviation, but indeed affects the extent of the initial inventory.

**6.3. The Effect of Quality Deterioration Rate on the Markdown Strategies.** The most valued characteristic of perishables is the perishability rate. The higher the perishability

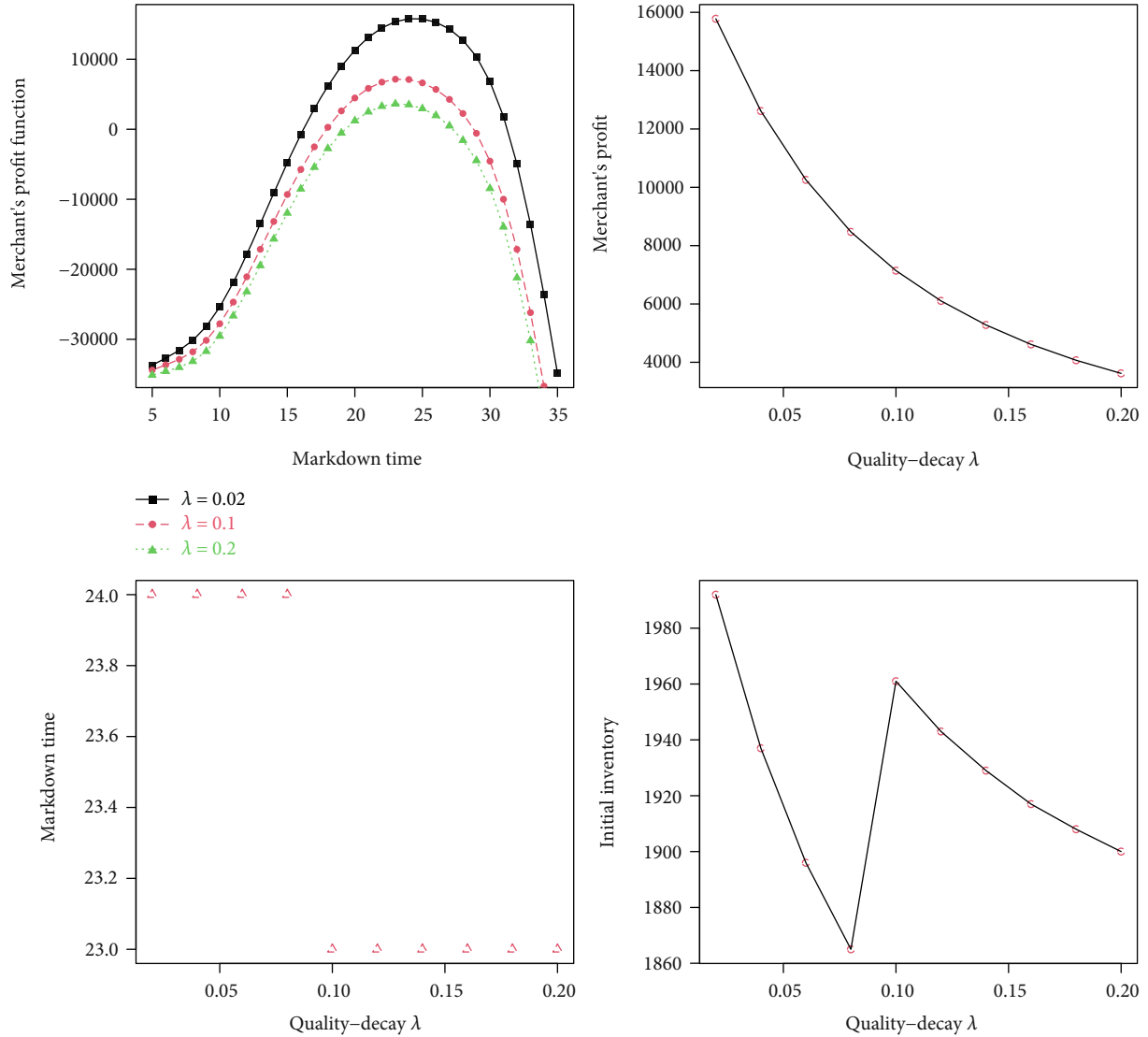


FIGURE 3: Effects' deterioration rate on markdown times.

rate, the more difficult it is to preserve. Figure 3 suggests that the merchant's profit decreases with the deterioration rate increase. So, the more perishable the product, the less profitable the merchant gets. The markdown time also does not seem too sensitive to the deterioration rate. It plays a role in regulating the change of the initial inventory to the deterioration rate. Generally, the shorter the shelf life of the product, the more difficult it is for the merchant to handle it, and the lower the final profit would be. In reality, some merchants operating in perishables always adopt high initial prices to quickly cover costs. In particular, products that are too perishable are even sold directly in presales or a custom model.

## 7. Conclusion

Freshness and price have become major issues in the management of perishables with the increasing attention paid to the freshness of perishables. The markdown strategy is a common operating method to reverse the declining demand

due to the dynamic quality evaluation. An effective markdown strategy can avoid waste due to the expiration dates when optimizing operational performance. Statistical big data based on advanced product identification and sensory technologies can capture perishables' freshness state and model the evolution of the freshness of perishables with shelf life (such as  $\beta q e^{-\lambda t}$ ) [45–49]. So, the time-sensitive markdown strategies are feasible. By introducing a time-variant market demand model, we consider the relationship between the optimal markdown time and the optimal initial inventory under uncertain market demand. The optimal markdown time as well as the initial inventory is derived under different scenarios. The results show that there exists an optimal markdown time under various scenarios; the uncertainty of market sales would force merchants to implement a markdown strategy in advance. The early implementation of the markdown strategy would then stimulate the potential market demand, which finally increases the merchant's profit. Hence, both the profit and initial inventory increase with



the uncertain market increase and decrease with the increase in the quality deterioration rate. The dynamic deterioration of perishable quality harms merchant's revenue. The more perishable the product, the less the merchant is willing to prepare in advance. Therefore, many perishables are sold in a customized way or presale in reality. Besides, markdown time is less sensitive to the uncertainty and perishability but has a regulatory effect on initial inventory.

Markdown time as an important issue in markdown strategy can provide new and significant insights for merchants. This research on time-sensitive markdown strategies enriches current markdown strategy research and provides a theoretical basis for an intelligent decision-making system for perishable product management. However, it should be noted that the markdown strategy in this study only focuses on the initial inventory and markdown time under uncertain market demand. It is undeniable that the initial price and the magnitude of the markdown price are also important factors in markdown strategies. Hence, future research on markdown strategies should include more decision variables to provide more insights into market strategies.

## Data Availability

The data used to support the findings of this study are available from the corresponding authors upon request.

## Conflicts of Interest

The authors have declared that no competing interests exist concerning this study.

## Acknowledgments

The work was supported by the Provincial Nature Science Foundation of Guangdong (2018A030313488, 2016A060311539), Guangdong Province Young Innovative Talents Program in Universities (Natural Science) (No. 2018KQNCX327), Guangdong Province Philosophy and Social Sciences "13th Five-Year" Planning Project for Building Discipline (GD17XGL56), Guangdong Soft Science Research Project (2018A070712005), and Fundamental Research Funds for the Central Universities (2018MSXM01).

## Supplementary Materials

The proofs of some theorems and properties are presented in the supplementary materials. (*Supplementary Materials*)

## References

- [1] S. Péneau, P. B. Brockhoff, F. Escher, and J. Nuessli, "A comprehensive approach to evaluate the freshness of strawberries and carrots," *Postharvest Biology and Technology*, vol. 45, no. 1, pp. 20–29, 2007.
- [2] J. Zhang, G. Liu, Q. Zhang, and Z. Bai, "Coordinating a supply chain for deteriorating items with a revenue sharing and cooperative investment contract," *Omega*, vol. 56, no. 3, pp. 37–49, 2015.
- [3] A. Şen, "Competitive markdown timing for perishable and substitutable products," *Omega*, vol. 64, pp. 24–41, 2016.
- [4] Y. Jiang, X. Gu, D. Wu et al., "A novel negative-transfer-resistant fuzzy clustering model with a shared cross-domain transfer latent space and its application to brain CT image segmentation," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 99, pp. 1–13, 2019.
- [5] Y. Jiang, Y. Zhang, C. Lin, D. Wu, and C. T. Lin, "EEG-based driver drowsiness estimation using an online multi-view and transfer TSK fuzzy system," *IEEE Transactions on Intelligent Transportation Systems*, vol. 99, pp. 1–13, 2020.
- [6] Y. Zhang, F. L. Chung, and S. Wang, "Clustering by transmission learning from data density to label manifold with statistical diffusion," *Knowledge-Based Systems*, vol. 193, p. 105330, 2019.
- [7] Y. Zhang, S. Wang, K. Xia, Y. Jiang, and P. Qian, "Alzheimer's disease multiclass diagnosis via multimodal neuroimaging embedding feature selection and fusion," *Information Fusion*, vol. 66, pp. 170–183, 2021.
- [8] Y. Zhang, J. Li, X. Zhou, T. Zhou, and J. Yang, "A viewreduction based multi-view TSK fuzzy system and its application for textile color classification," *Journal of Ambient Intelligence and Humanized Computing*, vol. 1, pp. 1–11, 2019.
- [9] T. Hsieh and C. Dye, "A production–inventory model incorporating the effect of preservation technology investment when demand is fluctuating with time," *Journal of Computational and Applied Mathematics*, vol. 239, no. 2, pp. 25–36, 2013.
- [10] P. H. Hsu, H. M. Wee, and H. M. Teng, "Preservation technology investment for deteriorating inventory," *International Journal of Production Economics*, vol. 124, no. 2, pp. 388–394, 2010.
- [11] Y. P. Lee and C. Dye, "An inventory model for deteriorating items under stock-dependent demand and controllable deterioration rate," *Computers and Industrial Engineering*, vol. 63, no. 2, pp. 474–482, 2012.
- [12] W. Elmaghraby and P. Keskinocak, "Dynamic pricing in the presence of inventory considerations: research overview, current practices, and future directions," *Management Science*, vol. 49, no. 10, pp. 1287–1309, 2003.
- [13] I. Hassouneh, T. Serra, and Š. Bojnec, "Nonlinearities in the Slovenian apple price transmission," *British Food Journal*, vol. 117, no. 1, pp. 461–478, 2015.
- [14] X. Wang and D. Li, "A dynamic product quality evaluation based pricing model for perishable food supply chains," *Omega*, vol. 40, no. 6, pp. 906–917, 2012.
- [15] I. Z. A. S. Karaesmen, *Managing Perishable and Aging Inventories: Review and Future Research Directions*, K. G. A. K. Kempf, Ed., Springer US, New York, NY, 2011.
- [16] N. H. Shah, D. B. Shah, and D. G. Patel, "Optimal preservation technology investment, retail price and ordering policies for deteriorating items under trended demand and two level trade credit financing," *Journal of Mathematical Modelling & Algorithms in Operations Research*, vol. 14, no. 1, pp. 1–12, 2015.
- [17] S. K. Goyal and B. C. Giri, "Recent trends in modeling of deteriorating inventory," *European Journal of Operational Research*, vol. 134, no. 1, pp. 1–16, 2001.
- [18] F. Raafat, "Survey of literature on continuously deteriorating inventory models," *The Journal of the Operational Research Society*, vol. 42, no. 1, pp. 27–37, 1991.

- [19] P. L. Abad, "Optimal pricing and lot-sizing under conditions of perishability and partial backordering," *Management Science*, vol. 42, no. 8, pp. 1093–1104, 1996.
- [20] P. K. Banerjee and T. R. Turner, "A flexible model for the pricing of perishable assets," *Omega*, vol. 40, no. 5, pp. 533–540, 2012.
- [21] A. Rajan and S. R. Rakesh, "Dynamic pricing and ordering decisions by a monopolist," *Management Science*, vol. 38, no. 2, pp. 240–262, 1992.
- [22] R. V. Ramasesh, "Lot-sizing decisions under limited-time price incentives: a review," *Omega*, vol. 38, no. 3-4, pp. 118–135, 2010.
- [23] Y. Levin, J. McGill, and M. Nediak, "Dynamic pricing in the presence of strategic consumers and oligopolistic competition," *Management Science*, vol. 55, no. 1, pp. 32–46, 2009.
- [24] C. H. Lee, T. M. Choi, and T. Cheng, "Selling to strategic and loss-averse consumers: stocking, procurement, and product design policies," *Naval Research Logistics (NRL)*, vol. 62, no. 6, pp. 435–453, 2015.
- [25] O. Perdikaki and J. Swaminathan, "Improving valuation under consumer search: implications for pricing and profits," *Production and Operations Management*, vol. 22, no. 4, pp. 857–874, 2013.
- [26] M. Yu, R. Kapuscinski, and H. Ahn, "Advance selling: effects of interdependent consumer valuations and seller's capacity," *Management Science*, vol. 61, no. 9, pp. 2100–2117, 2015.
- [27] Y. Aviv and A. Pazgal, "Optimal pricing of seasonal products in the presence of forward-looking consumers," *Manufacturing & Service Operations Management*, vol. 10, no. 3, pp. 339–359, 2008.
- [28] E. Adida and O. Ozer, "Why markdown as a pricing modality?," *Management Science*, vol. 65, no. 5, pp. 2161–2178, 2019.
- [29] G. Gallego, R. Phillips, and Ö. Şahin, "Strategic management of distressed inventory," *Production and Operations Management*, vol. 17, no. 4, pp. 402–415, 2010.
- [30] G. P. Cachon and P. Feldman, "Price commitments with strategic consumers: why it can be optimal to discount more frequently... than optimal," *Manufacturing & Service Operations Management*, vol. 17, no. 3, pp. 399–410, 2015.
- [31] Ö. Özer and Y. Zheng, "Markdown or everyday low price? The role of behavioral motives," *Management Science*, vol. 62, no. 2, pp. 326–346, 2015.
- [32] T. Choi and C. Chiu, Eds., *Risk Analysis in Stochastic Supply Chains: A Mean-Risk Approach*, Springer Science & Business Media, New York, 2012.
- [33] Z. Wu, W. Zhu, and P. Crama, "The newsvendor problem with advertising revenue," *Manufacturing & Service Operations Management*, vol. 13, no. 3, pp. 281–296, 2011.
- [34] G. P. Cachon and R. Swinney, "The value of fast fashion: quick response, enhanced design, and strategic consumer behavior," *Management Science*, vol. 57, no. 4, pp. 778–795, 2011.
- [35] M. S. Chen and C. C. Chuang, "An extended newsboy problem with shortage-level constraints," *International Journal of Production Economics*, vol. 67, no. 3, pp. 269–277, 2000.
- [36] Y. Li, F. Ye, and Q. Lin, "Optimal lead time policy for short life cycle products under conditional value-at-risk criterion," *Computers and Industrial Engineering*, vol. 88, no. 88, pp. 354–365, 2015.
- [37] Y. Qin, J. Wang, and C. Wei, "Joint pricing and inventory control for fresh produce and foods with quality and physical quantity deteriorating simultaneously," *International Journal of Production Economics*, vol. 152, pp. 42–48, 2014.
- [38] Y. Qin, R. Wang, A. J. Vakharia, Y. Chen, and M. M. H. Seref, "The newsvendor problem: review and directions for future research," *European Journal of Operational Research*, vol. 213, no. 2, pp. 361–374, 2011.
- [39] J. Wang, L. Wang, F. Ye, X. Xu, and J. Yu, "Order decision making based on different statement strategies under stochastic market demand," *Journal of Systems Science and Systems Engineering*, vol. 22, no. 2, pp. 171–190, 2013.
- [40] J. Wang, S. Zhang, C. Wu, J. Yu, C. Chen, and S. Tsai, "Time-sensitive markdown strategies for perishable products based on dynamic quality evaluation," *Kybernetes*, vol. 3, no. 7, pp. 1–16, 2020.
- [41] P. S. Taoukis and T. P. Labuza, "Applicability of time-temperature indicators as shelf life monitors of food products," *Journal of Food Science*, vol. 54, no. 4, pp. 783–788, 1989.
- [42] G. Fibich, A. Gaviols, and O. Lowengart, "Explicit solutions of optimization models and differential games with nonsmooth (asymmetric) reference-price effects," *Operations Research*, vol. 51, no. 5, pp. 721–734, 2003.
- [43] R. Sayner, "Markups & markdowns," *Independent Retailer*, p. 1, 2011.
- [44] G. Gallego and G. Van Ryzin, "Optimal dynamic pricing of inventories with stochastic demand over finite horizons," *Management Science*, vol. 40, no. 8, pp. 999–1020, 1994.
- [45] G. Khatwani and P. R. Srivastava, "Impact of information technology on information search channel selection for consumers," *Journal of Organizational and End User Computing*, vol. 30, no. 3, pp. 63–80, 2018.
- [46] N. Suyel and R. Pinki, "PpBAC: popularity based access control model for cloud computing," *Journal of Organizational & End User Computing*, vol. 30, no. 4, pp. 14–31, 2018.
- [47] L. Fabisiak, "Web service usability analysis based on user preferences," *Journal of Organizational and End User Computing (JOEUC)*, vol. 30, no. 4, pp. 1–13, 2018.
- [48] J. Jakli, T. Grubljei, and P. S. Coelho, "The shift to socio-organizational drivers of business intelligence and analytics acceptance," *Journal of Organizational and End User Computing (JOEUC)*, vol. 31, no. 2, pp. 37–64, 2017.
- [49] A. Shahri, M. Hosseini, K. Phalp, J. Taylor, and R. Ali, "How to engineer gamification: the consensus, the best practice and the grey areas," *Journal of Organizational and End User Computing*, vol. 31, no. 1, pp. 39–60, 2019.

## Research Article

# A Fast Hybrid Strategy-Based RFID Tag Identification Protocol

**Xinyan Wang** 

*Intelligent Science and Information Engineering College, Xi'an Peihua University, Xi'an 710125, China*

Correspondence should be addressed to Xinyan Wang; 36404843@qq.com

Received 22 December 2020; Revised 9 January 2021; Accepted 19 January 2021; Published 3 February 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Xinyan Wang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Tag collision is one of the critical problems in radiofrequency identification (RFID) technology which can be widely used to identify objects using tag attachment automatically. Through the transmission and reflection of wireless radiofrequency signals, noncontact identification is realized. However, when multiple tags respond to the reader simultaneously, a collision occurs, significantly degrading the identification performance of RFID systems. To tackle the tag collisions, we propose a fast hybrid strategy-based RFID anticollision (FHS-RAC) protocol. Based on the conventional query tree algorithm, the proposed FHS-RAC makes full use of collision bits and the total response bits to achieve the faster tag identification. Extensive simulations and experiments verify the feasibility and effectiveness of our proposed scheme.

## 1. Introduction

As the development of radiofrequency identification (RFID) technology, the RFID tag has become more and more inexpensive, facilitating a wide deployment of RFID systems. Meanwhile, the advent of low-cost RFID reader modules has made feasible the integration of RFID readers and mobile devices [1]. The direct adaptability of RFID to the demands of supply chains [2–4], object tracking and monitoring [5–7], traceability of patients, or the management of medication for elderly people are a number of applications that have, in recent years, increased the popularity of this technology. In these broader range of RFID-enabled applications, tag identification problem [8] is one of crucial issues, which is to allow the unread tagged objects appear in the interrogation area to be identified as fast as possible. Tag identification problem happens when unidentified tagged objects are moved in or the tagged objects are misplaced in the region of other readers. For example, in a shopping mall with thousands of tagged objects, a lot of new tagged commodities are transported into the warehouse or placed on the shelves every day. Often, RFID systems suffer from collisions when a reader identifies multiple tags or multiple readers contact with a tag simultaneously. As a result, tag reading failure occurs and tag reading efficiency degrades significantly. In order to solve the tag identification problem in a dense envi-

ronment, an anticollision mechanism is needed to coordinate the communication between the reader and tags, which is named the tag identification or anticollision algorithm.

To address tag collisions, a large number of anticollision protocols have been proposed, which can be divided into probabilistic [9–14], deterministic [15–18], and hybrid algorithms. The probabilistic algorithms are essentially based on the ALOHA protocol. The basic idea behind this kind of algorithm is that the reader estimates the number of tags according to the collision, success, and idle time slots first and then adjusts the frame length dynamically based on the estimation results. When the number of tags is equal to the number of time slots, the ALOHA-based algorithm can achieve the highest throughput rate of 36.8%. Ease of implementation is an advantage of the algorithm based on ALOHA. The randomness of the identification process and the starvation problem are the disadvantages of such algorithms. Deterministic algorithms are studied in articles [15–18]. The basic idea behind this algorithm is that first using bit identification and tracking technology to identify the collision bit accurately. The tags are then grouped according to the collision bits until there is only one tag in the group. One disadvantage of these protocols is that the length of probe command of such algorithm is not fixed; it will change with the change of the query prefix length during the tag identification process. In addition, query tree-based algorithm consumes more transmitted bits.

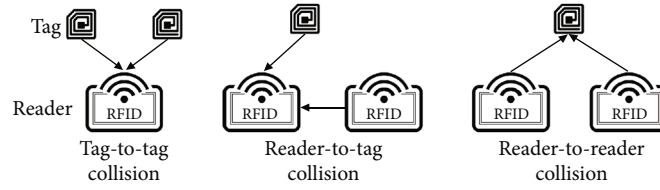


FIGURE 1: Three types of RFID collisions.

Combining the advantages of the ALOHA-based algorithm and the deterministic algorithm, the literatures [19, 20] proposed the hybrid-based algorithm. In the paper [20], based on the dynamic frame slot ALOHA algorithm, the binary tree algorithm is used to identify the collision slot. Compared with the ALOHA algorithm, the hybrid-based algorithm has a more considerable improvement in the throughput rate. However, the starvation problem is still to be solved, and its implementation complexity is higher than the ALOHA-based algorithm.

Considering the stable nature of deterministic algorithm, we propose a fast hybrid strategy-based RFID anticollision (FHS-RAC) with adaptive variable fork number. The proposed algorithm increases the number of probe forks based on the common query tree protocol, so that it can speed up the tag identification process. To avoid the problem that the reduced efficiency caused by the extra empty slots, the collision bits and the total response bits are used to estimate the tag cardinality, and the number of probe forks is adaptively adjusted according to the continuous situation of the highest collision bit. The proposed FHS-RAC can reduce the empty slots and quicken the convergence of query progress. Moreover, its efficiency will not be affected by the number of tags and tag ID distribution.

The remainder of this paper is organized as follows. In Section 2, we reviewed the collision types and the corresponding anticollision algorithms. Section 3 presents the system model and our proposed hybrid algorithm. In Section 4, mathematical analysis and performance evaluation through simulations is conducted. Section 5 concludes the whole paper.

## 2. Related Works

**2.1. Static Tag Identification.** To improve tag reading efficiency, many efforts have been made to solve RFID collision problems, which can be generally classified into three categories: tag-to-tag collisions, reader-to-tag collisions, and reader-to-reader collisions, as illustrated in Figure 1. Specifically, when one reader attempts to identify multiple tags, the tag responses reach the reader simultaneously and collide with each other (i.e., tag-to-tag collisions); when multiple readers try to read a tag, the reading request sent by one reader collide with the tag response to another reader (i.e., reader-to-tag collisions) or the signals transmitted by multiple readers may collide at the tag (i.e., reader-to-reader collisions).

Numerous tag anticollision algorithms have been proposed to reduce tag-to-tag collisions. Generally, there are two types of tag anticollision algorithms: ALOHA-based

algorithms and tree-based algorithms. The basic idea of ALOHA-based algorithms is that a tag waits for a random period of time and then retransmits data when collisions are detected. MAP [10] is an improved ALOHA-based algorithm for reducing tag reading collisions in the Internet of Things. Such an algorithm performs well, especially when the number of tags is small. In [14], an enhanced ALOHA-based algorithm has been presented to handle tag-to-tag collisions by querying the collided slots and then using different methods to adjust the frame length for collision resolution. However, because ALOHA-based algorithms randomly schedule data transmissions, some tags cannot be identified for a long time, causing the well-known tag starvation problem [16]. To address this problem, tree-based algorithms have been proposed and extensively studied [8, 15–17]. The principle of tree-based algorithms is to recursively divide tags into smaller subsets until each subset has only one or zero tags. In [8], an efficient collision arbitration algorithm, based on the binary tree algorithm, was designed to improve the performance of RFID systems through empty slot skipping. GBSA [17] is a tree-based collision arbitration strategy that integrates an efficient tag cardinality estimation method, an optimal grouping strategy, and a modified binary splitting. Although tree-based algorithms can eliminate tag starvation in dense tag scenarios, they often require lots of interactive processes, resulting in low reading efficiency [18].

To minimize reader-to-reader and reader-to-tag collisions, many reader anticollision protocols have been designed. Basically, they can be classified into two categories: distributed protocols and centralized protocols. As a distributed protocol, DCS [21] employs time division multiple access (TDMA) and divides the time into frames, each containing several time slots (colors) for tag readings. On top of this framework, an algorithm was designed to ensure different readers to choose different colors for collision avoidance. PDCS [22] is an enhanced version of DCS by enabling two collided readers to reselect their colors with a certain probability if they happen to choose the same color. EDCCS [23], another enhancement of DCS, can increase the number of contemporary readers by allowing collided readers to complete tag identification in prior time slots. The above algorithms assume static or indoor scenarios and do not consider readers' mobility. Thus, they are not suitable for mobile tag identification.

**2.2. Mobile Tag Identification.** Recently, some research has been conducted to ensure tag reading performance in mobile environments. In [20], the authors proposed an anticollision protocol, called dynamic collision tree protocol, to identify many RFID tags that dynamically move into and leave the

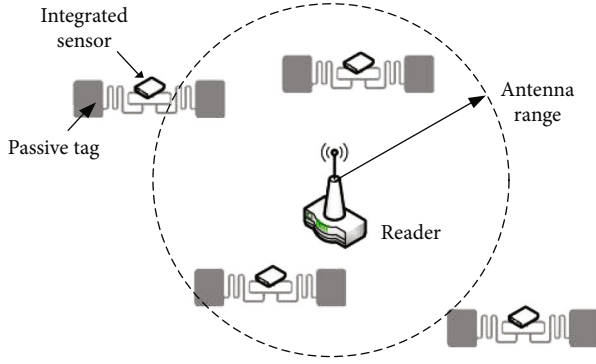


FIGURE 2: A scenario of RFID sensor network.

reading area in mobile RFID systems. MTIP [24] presents a feasible mobile tag estimation model, in which the superposition principle of Poisson process is used to solve the tag estimation problem. Moreover, an anticollision algorithm was proposed for identifying mobile RFID tags. To accelerate the reading process of large-scale mobile RFID systems, EBD [25] was designed to rapidly and efficiently distinguish and identify the unknown tags (i.e., not previously recognized tags) from the known tags while they are passing a reader's range. All these schemes focus on the situations in which many tags move through the reading area of a reader; thus, they can only handle tag-to-tag collisions. In contrast, this paper targets road environments where multiple vehicles (readers) pass a series of nonoverlapping tags sequentially. Hence, reader-to-tag and reader-to-reader collisions are the major issues to be addressed. Also, onboard tag readings are more challenging due to higher mobility and lower reliability.

### 3. System Model and Algorithm Description

**3.1. System Model.** A standard UHF RFID system, as shown in Figure 2, consists of at least one interrogator (reader) and a minimum of one transponder (tag). Such platforms communicate in the industrial, scientific, and medical (ISM) band around 900 MHz using the EPC Class 1 Generation 2 (EPC-C1G2) communication standard. Depending on the region of operation, the frequency bands and regulations differ (e.g., EU...865-868 MHz; NA...902-928 MHz) [2]. UHF RFID tags can be grouped depending on the technology incorporated. In this paper, we consider a large-scale RFID system with a single reader and a large number of EPC-C1G2 tags. To read a tag, the reader first transmits a reading request (RREQ). Then, the tag within the reading range collects and stores the energy carried by the signals. Once the energy reaches a certain threshold, the tag sends out a reading reply (RREP). As mentioned in Section 2.1, such communication process may suffer from three types of collisions. In this paper, we exclusively focus on the tag-to-tag collisions.

**3.2. Communication Slots.** The identification of FHS-RAC algorithm general includes three types of time slots: (1) singleton slot: that is picked by exactly one tag; (2) collision slot:

that is picked by multiple tags; and (3) empty slot: that is picked by none of tags. The total number of slots required for a multifork tree is

$$\begin{cases} t(m) = 1 + M \sum_{L=0}^{\infty} M^L K, \\ K = 1 - (1 - M^{-L})^m - m M^{-L} (1 - M^{-L})^{m-1}, \end{cases} \quad (1)$$

where  $M$  is the number of branches of the tree,  $m$  is the number of tags, and  $L$  is the current search depth. We know that when  $M$  takes different values, the total number of time slots consumed by the FHS-RAC is different. For example, when  $M$  is 3, the total number of time slots is the smallest, that is, the tag identification using 3-ary is the most efficient. However, it is impossible to construct a 3-ary in the practical identification process. In the actual process, the most common is to build 2-ary, 4-ary, or 8-ary traversal tree. Some algorithms use a 4-ary or 8-ary to reduce the collisions. However, the multitree will introduce a lot of extra empty time slots, it is necessary to prune the branch of multitree and thus to reduce the empty slots.

**3.3. Algorithm Description.** In this paper, the workflow of our proposed FHS-RAC algorithm can be described as follows. In FHS-RAC, the reader will calculate the collision factor and determine the current search depth and then adaptively adjust a 2-ary tree to a 4-ary tree or 8-ary tree to avoid the generation of overmuch empty slots, where the search depth is defined as the number of branches from the root to the leaf during the traversal tree, which is also the number of layer of the multifork tree. After dynamically selecting the appropriate  $M$ -ary tree, the collision bits are extracted to form a new query prefix. When receiving the query prefix, the tag compares its own ID with it and locks its own bits to reduce the amount of messages to be sent. In our proposed FHS-RAC, it is not necessary to increase the number of fork tree in order to reduce the collisions, because the number of empty slots also needs to be considered at the same time. Obviously, the higher the fork tree, the fewer collision slots and the more empty slots. Therefore, we need to adaptively adjust the fork tree to reduce collisions while limiting the increase of empty slots. In what follows, we describe the adjustment mechanism. In this paper, we define a variable called collision factor, which is expressed as

$$\mu = \frac{\alpha}{\beta}, \quad (2)$$

where  $\alpha$  is the number of collision bits and  $\beta$  is the length of tag response bits in a collision slot. Suppose there are  $m$  tags to be identified in the RFID system that meet the query conditions. The probability that any bit does not collide is  $1/2^{m-1}$ . Then, the collision factor  $\mu$  can be reexpressed as

$$\mu = \frac{1 - (1/2)^{m-1}}{\beta} = 1 - \left(\frac{1}{2}\right)^{m-1}. \quad (3)$$

From Equation (3), we know that the larger the number of concurrent tags, the higher the collision factor, and vice versa. Suppose the number of multitree assigned by the system is  $L$ , when the search depth is 1, the identification probability of a tag is

$$P_1 = \left(1 - \frac{1}{L}\right)^{m-1}, \quad (4)$$

when the search depth is  $k$ , the corresponding identification probability is

$$P_k = P_1(1 - P_1)^{k-1}. \quad (5)$$

The expected value of the search depth can be calculated as

$$E(k) = \sum_{k=1}^{\infty} k \times P_k = \sum_{k=1}^{\infty} k \times P_1 \times (1 - P_1)^{k-1}. \quad (6)$$

The above formula can be simplified as

$$E(k) = \frac{1}{(1 - (1/L))^{m-1}}. \quad (7)$$

Thus, the average number of time slots of a 2-ary tree is

$$T_2 = \frac{2}{(1 - 1/2)^{m-1}}. \quad (8)$$

Similarly, the average number of time slots of a 4-ary tree is

$$T_4 = \frac{4}{(1 - 1/4)^{m-1}}. \quad (9)$$

Comparing (2) to (4), we can get the following judgment conditions:

$$\begin{cases} T_2 \geq T_4, m \geq 3, \\ T_2 < T_4, m < 3. \end{cases} \quad (10)$$

Thus, by judging the threshold of the collision factor, the reader can adaptively adjust the multitree to resolve the collided tags. When the value of  $\mu$  is equal to or greater than 0.75, a 4-ary tree is used. And when the value of  $\mu$  is less than 0.75, a 2-ary tree is used. The collision factor is an indicator that reflects the number of tags in the collision slot. However, such indicator is not completely reliable because it is easily affected by the distribution of tag IDs. Therefore, we use the collision factor as the threshold and the highest collision bit to decide whether to adjust the fork tree. The workflow of the proposed FHS-RAC is described as follows:

- (S1) The reader maintains an empty query stack \$\$\$\$. First, the reader broadcasts a request, and the tags

within the reader's coverage receive the request command and make the corresponding operations

- (S2) The reader determines the real-time status of slots according to the tag response. If only one tag replies, the reader successfully identifies the tag and jumps to S5; if no tag replies, the search is stopped in the branch, jump to S5; if more than one tag reply, the reader determines the tags collide and jump to S3
- (S3) The reader determines whether the highest collision bit is continuous. If the highest collision bit is not continuous, the reader directly selects the 2-ary tree. If the highest collision bit is continuous, the reader skips to S4
- (S4) The reader calculates the collision factor  $\mu$ . If  $\mu \leq 0.75$ , the reader selects a 2-ary tree. If  $\mu > 0.75$ , the reader selects a 4-ary tree
- (S5) Judging whether the stack is empty, if not, the reader extracts the first prefix of the query stack to probe tags, and returns to S3; otherwise, the entire identification process ends

## 4. Performance Analysis and Numerical Results

*4.1. Performance Analysis.* We theoretically analyze the total number of slots required for our proposed FHS-RAC algorithm and then deduce the system efficiency. Specifically, the total number of slots can be obtained by summing the number of slots consumed by multitree. The average number of tags contained in each intermediate node is 3 in  $k$ th search depth. In the tag identification process, if the required traversal depth is higher than  $k$ , the reader should use a 2-ary tree; otherwise, it should use a 4-ary tree. The total number of slots required by FHS-RAC can be expressed as

$$T_{\text{total}} = T_4 + T_2, \quad (11)$$

where  $T_4$  means the total number of slots taken by a full 4-ary tree, which is calculated as

$$T_4 = \sum_{i=0}^k 4^i = \sum_{i=0}^{\lceil \log_4^{M/3} \rceil - 1} 4^i. \quad (12)$$

Referring to the analysis in [18], the number of total slots ( $T_{4\text{-total}}$ ), empty slots ( $T_{4\text{-idle}}$ ), and collision slots ( $T_{4\text{-coll}}$ ) can be calculated as

$$\begin{aligned} T_{4\text{-total}}(m) &= 1 + 4 \sum_{j=0}^{\infty} 4^j \left[ 1 - (1 - 4^{-j})^m - \frac{m}{4^j} (1 - 4^{-j})^{m-1} \right], \\ T_{4\text{-coll}} &= \frac{1}{4} (T_{4\text{-total}} - 1), \\ T_{4\text{-idle}} &= T_{4\text{-total}} - T_{4\text{-coll}} = \frac{3}{4} T_{4\text{-total}} - m + \frac{1}{4}, \end{aligned} \quad (13)$$

in which  $m$  denotes the number of tags involved in current slot and  $j$  represents the traversal depth. Referring to the existing literature [16], empty slots can be eliminated by the introduction of *QueryRP* command. Therefore, the total number of slots taken by FHS-RAC to identify  $n$  tags is expressed as

$$\begin{aligned} T_{\text{FHS-RAC}}(n) &= T_{4\text{-total}} + T_{2\text{-total}} - T_{4\text{-idle}} + T_{4\text{-coll}} \\ &\approx \sum_{j=0}^{\lceil \log_4^{n-1} \rceil - 1} 4^j + \frac{5}{3}(n-s) - 1.16(n-s) \\ &\quad + 0.72(n-s). \end{aligned} \quad (14)$$

Then, the system efficiency of FHS-RAC can be calculated as  $n/T_{\text{FHS-RAC}}$ .

**4.2. Numerical Results.** In this section, we evaluate the tag identification performance of our proposed algorithm and the existing algorithms DBSA, DFSA [10], and IACA [15]. We implement the proposed algorithm in MATLAB on a ThinkPad X1 Carbon desktop with an Intel 2.4GHz CPU. Our simulation setting follows the specifications of the EPC C1 Gen2 standard. The number of tags is from 20 to 500 in step of 20. The reader-to-tag transmission rate and the tag-to-reader data rate are not symmetric, which depends on specific physical implementations and practical environments. We analyze the performance of algorithms by using three metrics: the average request cycle and the total length of transmitted data between the reader and all the tags and the energy efficiency of the tags.

During our simulations, we randomly generated 16-bit ID tags and the number of tags is set between 5 and 500 tags. The average request cycle can be computed as

$$N_{\text{average}} = \frac{N_{\text{total}}}{n}, \quad (15)$$

herein,  $N_{\text{total}}$  is the total number of the request cycles which is the sum of the number collision, identification, and empty cycles. We denote  $E_{tr}$  and  $E_{tt}$  as the power consumption of tag receiving queries and transmitting responses, respectively. Thus, the energy efficiency can be written as

$$\eta = \frac{nE_u}{QE_{tr} + RE_{tt}}, \quad (16)$$

where  $Q$  is the number of the queries and  $R$  is the total number of tag responses.

Figure 3 shows the average cycles for one-tag identification as the number of tags is increased. Our proposed scheme takes 2.25 cycles, whereas DBSA and IACA take 4.90 and 2.98 cycles, respectively. Note that Figure 4 depicts the total length of transmitted data of DBSA, IACA, DFSA, and our proposed algorithm. The length of transmitted data of our proposed algorithm is significantly less than that in the DBSA and IACA algorithms. As the number of tags is increasing, the advantages of the proposed scheme become more noticeable, especially in the case of a very large number of tags. For example, when the number of tags is 405, the DBSA and IACA need to send a total length about 253756

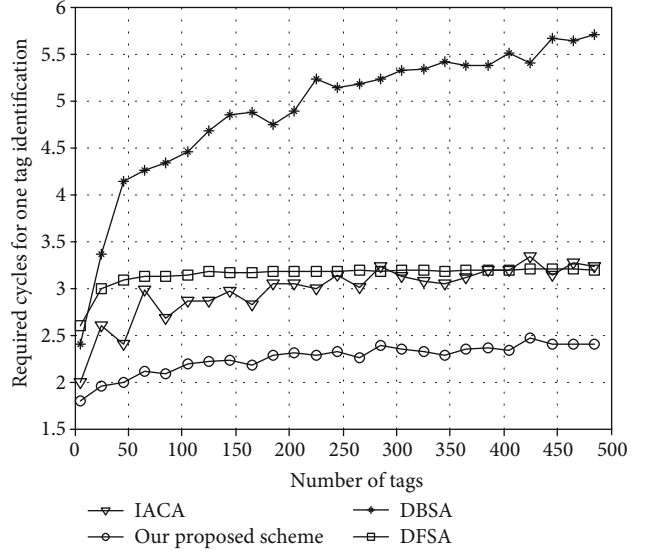


FIGURE 3: Required cycles for one tag identification.

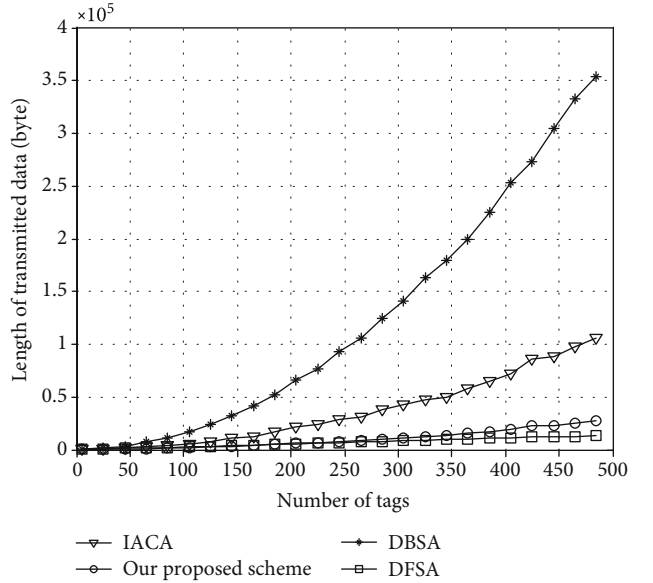


FIGURE 4: The length of transmitted data for all tag identifications.

and 77650 bytes, but our proposed algorithm needs to send only 19732 bytes. We found, in the results, the ALOHA-based algorithm outperforms tree-based algorithms in terms of the length of transmitted data.

Figure 5 shows the channel saving provided by our proposed algorithm compared to the DBSA and IACA, which are calculated by the following equation:

$$\begin{aligned} P_{\text{ChannelSaving}}^1 &= \left(1 - \frac{N_{\text{bits}}_{\text{IACA}}}{N_{\text{bits}}_{\text{DBSA}}}\right), \\ P_{\text{ChannelSaving}}^2 &= \left(1 - \frac{N_{\text{bits}}_{\text{OurScheme}}}{N_{\text{bits}}_{\text{DBSA}}}\right), \\ P_{\text{ChannelSaving}}^3 &= \left(1 - \frac{N_{\text{bits}}_{\text{OurScheme}}}{N_{\text{bits}}_{\text{IACA}}}\right). \end{aligned} \quad (17)$$

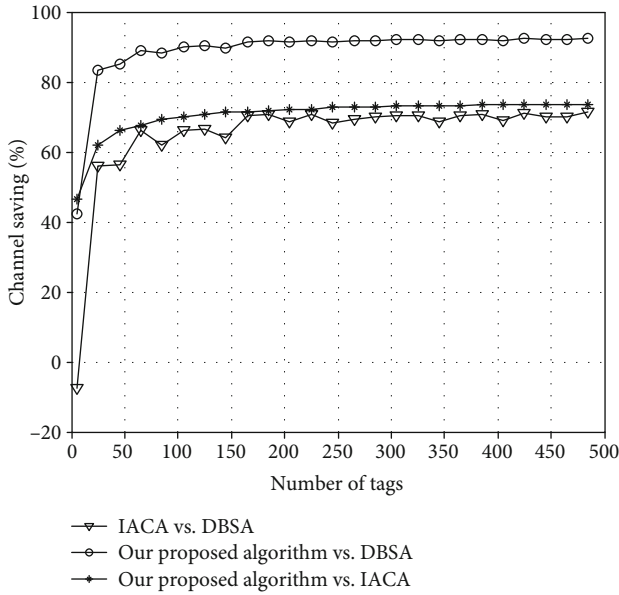


FIGURE 5: Channel saving under different algorithms.

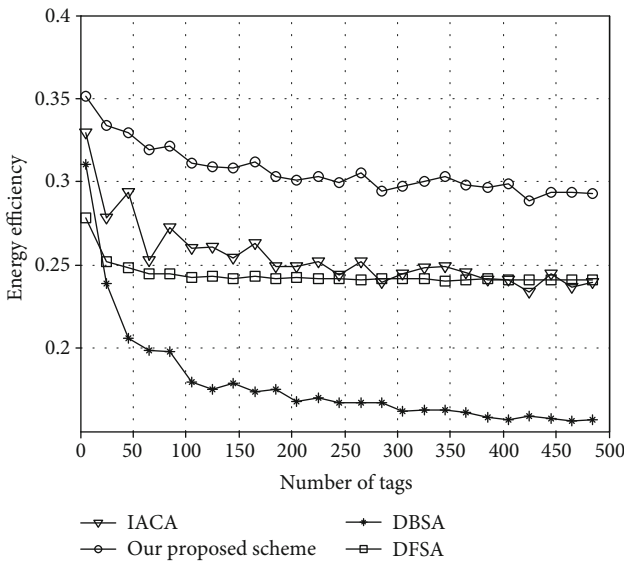


FIGURE 6: Comparison in energy efficiency.

From Figure 5, the results show that almost 70% channels are saved by our proposed algorithm than the IACA due to the use of additional request command's building and the mapping of first and second collision bits.

The comparison of energy efficiency is plotted in Figure 6. It is obvious that our proposed algorithm provides much higher energy efficiency than DBSA, IACA and DFSA.

## 5. Conclusion

This paper investigates the problem of tag collisions in the large RFID systems. We propose an efficient and complete a new tag identification algorithm, which is able to fast identify unread tags in a highly efficient and complete way. With

our algorithm, the request cycles and the length of transmitted data between the reader and tags are significantly reduced. The performance is achieved by the building of request command and the mapping of collision bits. Through simulation results, we conclude that our algorithm reduces time complexity, energy cost, and communication overhead in the identification process of multiple tags.

## Data Availability

To be frank, I derived the writing material from different journals as provided in the references. A MATLAB tool has been utilized to simulate our concept.

## Conflicts of Interest

The author declares that he has no conflicts of interest.

## References

- [1] S. S. Park, "An IoT application service using mobile RFID technology," in *2018 International Conference on Electronics, Information, and Communication (ICEIC)*, pp. 1–4, Honolulu, HI, USA, January 2018.
- [2] C.-H. Lee and C.-W. Chung, "RFID data processing in supply chain management using a path encoding scheme," *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 5, pp. 742–758, 2011.
- [3] J. Su, R. Xu, S. Yu, B. Wang, and J. Wang, "Idle slots skipped mechanism based tag identification algorithm with enhanced collision detection," *KSII Transactions on Internet and Information Systems*, vol. 14, no. 5, pp. 2294–2309, 2020.
- [4] J. Su, R. Xu, S. Yu, B. Wang, and J. Wang, "Redundant rule detection for software-defined networking," *KSII Transactions on Internet and Information Systems*, vol. 14, no. 6, pp. 2735–2751, 2020.
- [5] L. Shanguan, Z. Zhouy, X. Zheng, L. Yang, Y. Liu, and J. Han, "Shopminer: mining customer shopping behavior in physical clothing stores with COTS RFID devices," in *SenSys '15: Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems*, pp. 113–126, New York, NY, USA, November 2015.
- [6] L. Yang, J. Cao, W. Zhu, and S. Tang, "Accurate and efficient object tracking based on passive RFID," *IEEE Transactions on Mobile Computing*, vol. 14, no. 11, pp. 2188–2200, 2015.
- [7] H. Cai, L. D. Xu, B. Xu, C. Xie, S. Qin, and L. Jiang, "IoT-based configurable information service platform for product lifecycle management," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 2, pp. 1558–1567, 2014.
- [8] J. Su, Z. Sheng, L. Xie, and G. Wen, "Idle-slots elimination based binary splitting anti-collision algorithm for RFID," *IEEE Communications Letters*, vol. 20, no. 12, pp. 2394–2397, 2016.
- [9] L. Zhu and T. S. Yum, "A critical survey and analysis of RFID anti-collision mechanisms," *IEEE Communications Magazine*, vol. 49, no. 5, pp. 214–221, 2011.
- [10] W. T. Chen, "An accurate tag estimate method for improving the performance of an RFID anticollision algorithm based on dynamic frame length ALOHA," *IEEE Transactions on Automation Science and Engineering*, vol. 6, no. 1, pp. 9–15, 2009.



- [11] J. Su, A. X. Liu, Z. Sheng, and Y. Chen, "A partitioning approach to RFID identification," *IEEE/ACM Transactions on Networking*, vol. 28, no. 5, pp. 2160–2173, 2020.
- [12] J. Su, Z. Sheng, A. X. Liu, Z. Fu, and Y. Chen, "A time and energy saving based frame adjustment strategy (TES-FAS) tag identification algorithm for UHF RFID systems," *IEEE Transactions on Wireless Communications*, vol. 19, no. 5, pp. 2974–2986, 2020.
- [13] B. Li and J. Y. Wang, "Efficient anti-collision algorithm utilizing the capture effect for ISO 18000-6C RFID protocol," *IEEE Communications Letters*, vol. 15, no. 3, pp. 352–354, 2011.
- [14] H. Chen, K. Liu, C. Ma, Y. Han, and J. Su, "A novel time-aware frame adjustment strategy for RFID anti-collision," *Computers, Materials & Continua*, vol. 57, no. 2, pp. 195–204, 2018.
- [15] M. Djeddou, R. Khelladi, and M. Benssalah, "Improved RFID anti-collision algorithm," *AEU-International Journal of Electronics and Communications*, vol. 67, no. 3, pp. 256–262, 2013.
- [16] J. Su, Y. Chen, Z. Sheng, Z. Huang, and A. X. Liu, "From M-ary query to bit query: a new strategy for efficient large-scale RFID identification," *IEEE Transactions on Communications*, vol. 68, no. 4, pp. 2381–2393, 2020.
- [17] J. Su, Z. Sheng, A. X. Liu, Y. Han, and Y. Chen, "A group-based binary splitting algorithm for UHF RFID anti-collision systems," *IEEE Transactions on Communications*, vol. 68, no. 2, pp. 998–1012, 2020.
- [18] D. R. Hush and C. Wood, "Analysis of tree algorithms for RFID arbitration," in *Proceedings 1998 IEEE International Symposium on Information Theory (Cat. No.98CH36252)*, pp. 107–116, Cambridge, MA, USA, August 1998.
- [19] H. Wu, Y. Zeng, J. Feng, and Y. Gu, "Binary tree slotted ALOHA for passive RFID tag anti-collision," *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 1, pp. 19–31, 2012.
- [20] X. Jia, M. Bolic, Y. Feng, and Y. Gu, "An efficient dynamic anti-collision protocol for mobile RFID tags identification," *IEEE Communications Letters*, vol. 23, no. 4, pp. 620–623, 2019.
- [21] J. Waldrop, D. W. Engels, and S. E. Sarma, "Colorwave: a MAC for RFID reader networks," in *2003 IEEE Wireless Communications and Networking, 2003. WCNC 2003*, pp. 1701–1704, New Orleans, LA, USA, March 2003.
- [22] F. Gandino, R. Ferrero, B. Montrucchio, and M. Rebaudengo, "Introducing probability in RFID reader-to-reader anti-collision," in *2009 Eighth IEEE International Symposium on Network Computing and Applications*, pp. 250–257, Cambridge, MA, USA, July 2009.
- [23] N. Wang, G. Xie, and Y. Liu, "EDCS: an enhanced distributed color selection algorithm for anti-collision in RFID system," in *2017 3rd IEEE International Conference on Computer and Communications (ICCC)*, pp. 1031–1035, Chengdu, China, December 2017.
- [24] Y. Chen and Q. Feng, "A collision avoidance identification algorithm for mobile RFID device," *IEEE Transactions on Consumer Electronics*, vol. 65, no. 4, pp. 493–501, 2019.
- [25] L. Zhang, W. Xiang, and X. Tang, "An efficient bit-detecting protocol for continuous tag recognition in mobile RFID systems," *IEEE Transactions on Mobile Computing*, vol. 17, no. 3, pp. 503–516, 2018.